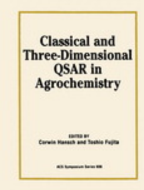


Classical and Three-Dimensional QSAR in Agrochemistry

EDITED BY
Corwin Hansch and Toshio Fujita

ACS Symposium Series [illegible]

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.fw001

ACS SYMPOSIUM SERIES **606**

Classical and Three-Dimensional QSAR in Agrochemistry

Corwin Hansch, EDITOR
Pomona College

Toshio Fujita, EDITOR
Fujitsu Kansai Systems Laboratory

Developed from a symposium sponsored
by the Division of Agrochemicals
at the 208th National Meeting
of the American Chemical Society,
Washington, DC,
August 21–25, 1994

American Chemical Society, Washington, DC 1995

S 583.2 .C58 1995 Copy 1

Classical and three-dimensional QSAR in agrochemistry

Library of Congress Cataloging-in-Publication Data

Classical and three-dimensional QSAR in agrochemistry / Corwin Hansch, editor, Toshio Fujita, editor.

p. cm.—(ACS symposium series; 606)

"Developed from a symposium sponsored by the Division of Agrochemicals at the 208th National Meeting of the American Chemical Society, Washington, DC, August 21–25, 1994."

Includes bibliographical references and index.

ISBN 0–8412–3321–7

1. Agricultural chemicals—Structure–activity relationships—Congresses.

I. Hansch, Corwin. II. Fujita, Toshio, 1929– . III. American Chemical Society. Division of Agrochemicals. IV. American Chemical Society. Meeting (208th: 1994: Washington, DC) V. Series.

S583.2.C58 1995
631.8—dc20 95–37029
CIP

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.fw001

This book is printed on acid-free, recycled paper.

Copyright © 1995

American Chemical Society

All Rights Reserved. The appearance of the code at the bottom of the first page of each chapter in this volume indicates the copyright owner's consent that reprographic copies of the chapter may be made for personal or internal use or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to copying or transmission by any means—graphic or electronic—for any other purpose, such as for general distribution, for advertising or promotional purposes, for creating a new collective work, for resale, or for information storage and retrieval systems. The copying fee for each chapter is indicated in the code at the bottom of the first page of the chapter.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

American Chemical Society
Library
1155 16th St., N.W.
Washington, D.C. 20036

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.fw001

1995 Advisory Board

ACS Symposium Series

Robert J. Alaimo
Procter & Gamble Pharmaceuticals

Mark Arnold
University of Iowa

David Baker
University of Tennessee

Arindam Bose
Pfizer Central Research

Robert F. Brady, Jr.
Naval Research Laboratory

Mary E. Castellion
ChemEdit Company

Margaret A. Cavanaugh
National Science Foundation

Arthur B. Ellis
University of Wisconsin at Madison

Gunda I. Georg
University of Kansas

Madeleine M. Joullie
University of Pennsylvania

Lawrence P. Klemann
Nabisco Foods Group

Douglas R. Lloyd
The University of Texas at Austin

Cynthia A. Maryanoff
R. W. Johnson Pharmaceutical
Research Institute

Roger A. Minear
University of Illinois
at Urbana–Champaign

Omkaram Nalamasu
AT&T Bell Laboratories

Vincent Pecoraro
University of Michigan

George W. Roberts
North Carolina State University

John R. Shapley
University of Illinois
at Urbana–Champaign

Douglas A. Smith
Concurrent Technologies Corporation

L. Somasundaram
DuPont

Michael D. Taylor
Parke-Davis Pharmaceutical Research

William C. Walker
DuPont

Peter Willett
University of Sheffield (England)

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.fw001

Foreword

THE ACS SYMPOSIUM SERIES was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of this series is to publish comprehensive books developed from symposia, which are usually "snapshots in time" of the current research being done on a topic, plus some review material on the topic. For this reason, it is necessary that the papers be published as quickly as possible.

Before a symposium-based book is put under contract, the proposed table of contents is reviewed for appropriateness to the topic and for comprehensiveness of the collection. Some papers are excluded at this point, and others are added to round out the scope of the volume. In addition, a draft of each paper is peer-reviewed prior to final acceptance or rejection. This anonymous review process is supervised by the organizer(s) of the symposium, who become the editor(s) of the book. The authors then revise their papers according to the recommendations of both the reviewers and the editors, prepare camera-ready copy, and submit the final papers to the editors, who check that all necessary revisions have been made.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previously published papers are not accepted.

Preface

THIS VOLUME STEMS FROM THE SYMPOSIUM honoring Toshio Fujita for his work in agricultural chemistry. The variety of the papers represent a few of the many areas of research in physical organic chemistry, toxicology, traditional and 3-D QSAR to which he has so effectively contributed. It is not only the quality but also the great diversity of his contributions which makes Toshio Fujita a leader in agricultural chemistry and many other aspects of QSAR.

As we point out in Chapter 1, QSAR has achieved its most notable commercial successes in the design of herbicides and insecticides. Little comparable to this is known from the area of medicinal chemistry. It is my belief that the exceptional success QSAR has achieved in agrochemistry is in large part due to the stimulation agricultural scientists received from the Kyoto University laboratory of Fujita, his colleagues, and his many students. His almost 300 papers analyze many aspects of chemico-biological interactions beyond those associated with agricultural chemistry.

I do not mean to imply that the only measure of success for QSAR is a commercial product. Indeed, in the last 30 years QSAR has changed our whole approach to the design of bioactive compounds and less toxic chemicals. No longer is it a great achievement to formulate an equation correlating chemical structure and biological efficacy. Excitement now comes from the quality of fit, the range of structures covered, mechanistic insight, and comparison with other QSAR.

The next phase in the QSAR adventure will be to better organize what we have learned so that past experience can be brought to bear on current research. This can be viewed as a complex extension of organic reactions. Once a new organic reaction is discovered, many researchers begin to delineate its scope and its side reactions (somewhat comparable to side effects and metabolism of bioactive compounds). Gradually these reactions are then related to the still growing body of organic chemistry.

QSAR appears to be evolving in a similar fashion. We are seeking out the similarities and differences in the way chemicals affect plants, animals, insects, cells, enzymes, etc. A compound affecting the biochemistry of one organism is likely to affect that of others. Fujita's EMIL project is an important innovative means for uncovering such generalizations. QSAR must go hand in hand with advances in comparative biochemistry and toxicology. To uncover the possibilities for selectively active

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.pr001

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

chemicals, QSAR will be increasingly important in forming the myriad connections between organic chemistry and biology.

It is my sincere hope that this volume will be helpful not only to agricultural scientists, but also to scientists working in related fields such as biochemistry, pharmaceuticals, and cosmetics.

CORWIN HANSCH
Department of Chemistry
Pomona College
645 North College Avenue
Claremont, CA 91711–6338

June 28, 1995

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.pr001

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Chapter 1

Status of QSAR at the End of the Twentieth Century

Corwin Hansch[1] and Toshio Fujita[2]

[1]Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711–6338
[2]EMIL Project, Fujitsu Kansai Systems Laboratory, 2–2–6 Shiromi, Chuo-ku, Osaka 540, Japan

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Despite the fact that the various approaches to QSAR have, during the past three decades, completely changed the way we study the interaction of organic compounds with the various forms of life, we have a long way to go in developing a science of QSAR. Some of the shortcomings, as well as the advantages, of our current methodologies are considered.

We are all caught up in the task of trying to improve different aspects of the QSAR paradigm. However, we need to consider from time to time the weakness as well as the advantages of the so-called classical QSAR, and of course the same applies to the cutting-edge developments. Weakness in classical QSAR arises from poor selection of substituents which can lead to collinearity problems, but can also result in structural changes which are not easily parameterized or cover parameter space poorly. Classical QSAR tends to be under-parameterized, especially in terms of steric effects. Its strength is that it provides the possibility for mechanistic interpretation of the SAR and lateral validation among QSAR. The Hammett-Taft parameters which are usually used in classical QSAR, are especially valuable in the comparison of biological QSAR with those for reactions of organic compounds in homogenous solutions. Until we can group QSAR in terms of mechanisms we are not developing a science of QSAR.

Classical QSAR

A little over three decades have passed since the discovery of a general approach for the formulation of biological quantitative structure activity relationships (QSAR) (*1,2*) The discovery fits in naturally with the subject of this symposium in that it grew out of the studies of plant growth regulators, especially the phenoxyacetic acids. Many have mistakenly thought that QSAR stems from medicinal chemistry. In fact, the first QSAR symposium, called "Biological Correlations -- The Hansch Approach," was sponsored by the ACS Division of Pesticide Chemistry (now Agrochemicals) at the 161st national meeting in Los Angeles in 1970.

0097–6156/95/0606–0001$12.00/0

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Classical QSAR and Agrochemistry

That the early start in QSAR occured in agrochemicals may account, in part, for more successes in the form of commercial pesticides than in the case of new drugs evolving from QSAR. QSAR played an important role in the design of the following commercial products: herbicides, Metamitrone (Bayer) (*3-5*), Bromobutide (Sumitomo) (*6*), the insecticide Bifenthrin (FMC) (*7*), and the fungicides Metconazole and Ipconazole (Kureka) (*8*). The present volume provides an opportunity for reflection on where the subject stands today.

The QSAR database being constructed at Pomona contains 6000 QSAR evenly divided between those from physical organic chemistry and those from biology (*9,10*). Among the biological QSAR, 109 are for insects, 67 for plants and 145 are for fungi (which are not limited to the agrochemical class). However, these figures understate the situation. For example, there are 88 enzymic QSAR from cholinesterase and 24 for P-450 enzymes used to develop insecticides and synergists. In addition there are 40 on chloroplasts which have been used to devise photosystem II inhibitors and 16 QSAR for algae. Many other QSAR on enzymes or organelles may also be of help in the design of bioactive compounds. Of course the total number of agrochemical QSAR is less than that for pharmaceuticals, but one should not forget, as the EMIL project has shown (*11,12*), that there is not a clean boundary in structural classes between compounds effective as pharmaceuticals or pesticides.

In the current database the largest number of QSAR (885) is for studies on cells (bacteria, tumor, etc.) followed by those for enzymes (773). The most interesting examples, of which many more are needed, are those where the same or similar compounds are tested on isolated enzymes and also on intact cells or whole organisms.

Postulates of Classical QSAR

The initial postulates, which still hold, are that there are three major factors needed to rationalize variations in a set of congeners producing a standard response in a test system (*2,13*): electronic, hydrophobic, and steric. It was the hope that if Hammett-Taft electronic parameters and logP or π hydrophobic parameters could account for two of these three major factors, then it might be possible to discern steric effects. These could be parameterized by Taft's E_S (especially for intramolecular effects), molar refractivity (essentially molar volume) or dimensional parameters such as STERIMOL. It is surprising that this simple set of principles has been used to elucidate literally thousands of QSAR for almost every imaginable type of chemico-biological interaction. Although more attention is now being given to H-bonding, it has not been a problem of overriding importance.

Quality of QSAR

We can now begin to obtain some general feeling for the quality of the results for classical QSAR. Table I has been formulated by searching the biological QSAR database with the command >. That is, **n > 7** finds the number of sets based on 8 or more data points (1984). Thus about one third of the examples are based on 7 or fewer points.

Table I. Distribution of QSAR Based on Set Size

Number of Congeners/Set	*Number of QSAR*	*Sets with Outliers*
n > 7	1984	1272
n > 12	1144	856
n > 20	568	450
n > 50	129	99
n > 100	31	25

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

While small sets cannot characterize much of receptor space, they can provide clues about electronic or hydrophobic effects which may be supported by lateral validation with other QSAR (*9,10,14,15*). Even a set of 100 congeners acting in an animal provides information for only a superficial characterization of the SAR. It is for this reason that lateral validation is so important. We must make the maximum use of all QSAR information, enzymes, P-450, organelles, membranes, etc., as well as that from physical organic chemistry in the design of new agrochemicals and drugs. Compounds in the set marked as outliers in a QSAR are not included in the correlation. The situation is not as bad as pictured in Table I. Based on 8 or more congeners (1984 sets) we find that 1272 contain one or more outliers. However, in many instances congeners lacking suitable parameters are held in the set, but are not used to formulate the equation. In another view of quality we can consider the sets where $n > 12$ (1144 examples) shown in Table II:

Table II. Number of QSAR out of 1144 at Three Levels of Quality

Level of Correlation	*Number of QSAR*	*Sets with Outliers (%)*
$r > .90, r^2 > .81$	832	576 (69)
$r > .95, r^2 > .90$	305	176 (58)
$r > .975, r^2 > .95$	107	56 (52)

The better correlations tend to have fewer outliers. Clearly a large number of equations leave 20% or more of the variance "unexplained." In many instances, especially in the early work, part of this can be charged to experimental noise, but there is plenty of room for improvement by better parameterization. Turning now to QSAR from the physical organic database we can compare these with the biological QSAR via Table III.

Table III. Distribution of Physical Organic QSAR Based on Set Size

Number of Congeners/Set	*Number of QSAR*	*Sets with Outliers (%)*
$n > 7$	1479	704 (48)
$n > 12$	572	316 (55)
$n > 20$	163	88 (54)
$n > 50$	17	71 (41)
$n > 100$	2	1 (50)

The data set size for physical organic chemistry is smaller at every level. Here too outliers in a set may be ortho substituents for which often no attempt was made at parameterization.

Table IV based on $n > 12$ (572 examples) for physical organic reactions can be compared with Table II.

Table IV. Number of Physical Organic QSAR at Three Levels

Level of Correlation	*Number of QSAR*	*Sets With Outliers (%)*
$r > .90, r^2 > .81$	538	288 (54)
$r > .95, r^2 > .90$	420	316 (75)
$r > .975, r^2 > 0.95$	272	121 (44)

The correlations are much better; 94% of the total number of examples have r greater than 0.90. Only 73% of the biological QSAR meet this standard. One can not expect as high quality for the biological QSAR in nonhomogenous systems where there are so many possibilities for side reactions, including metabolism. In fact, before 1962 it did not seem to be possible to derive QSAR even for the simple case of enzymes.

There is the feeling that outliers are omitted from the correlations simply to obtain a higher value of r^2. There are serious reasons for omitting data points. They may indicate that the parameterization is not correct. They may indicate that certain compounds are operating by a different mechanism or undergoing metabolism at a greater rate than the others in the set. Finally, we have found over the years that experimental errors in the data may be at fault.

We now compare results with enzymes (775 examples) with those from whole organisms (insects, plants, animals; 719 examples).

Table V. Quality of QSAR for Enzymes and Whole Organisms

	Level of Correlation	*Number of QSAR*	*% of Total*
Enzymic:	$r > .90, r^2 > .81$	648	84%
	$r > .95, r^2 > .90$	373	48%
Whole Organism:	$r > .90, r^2 > .81$	535	74%
	$r > .95, r^2 > .90$	285	40%

The disparity in quality is not as great as one might have expected.

Collinearity Problem and Substituent Selection

The paucity of truly well designed studies is readily apparent to the careful reader of the literature. Possibly the greatest shortcoming is the haphazard way in which structural changes are made in the game of "molecular roulette." These changes, based on hunches, often preclude any attempt at QSAR. Even today the importance of the collinearity problem is by no means universally understood. In the first set, one needs to maximize variance and minimize collinearity of substituent changes which can be well parameterized.

In the design of an evolving study, close cooperation between three types of expertise is urgent: synthetic chemistry, biological testing and QSAR formulation. In the beginning, synthesis should produce a data set amenable to good parameterization. Unless synthesis is very difficult, this should provide a starter set from which a few SAR ideas can be gleaned from the QSAR. A good idea might be to first place an easily introduced substituent (e.g. Cl or CH_3) at each easily substituted point on the parent molecule. At those positions which show a response further devlopement can be made. B. R. Baker called this prospecting for bulk tolerance (*16*).

Another important problem in substituent selection is that of avoiding easily metabolized substituents. Esters fall in this class and should not be used in the early stages of a study. Nitro compounds are easily reduced in vivo. In place of an ester one might use the more stabile acyl group and a cyano group is a good stand-in electronically for a nitro group. In a rough general way metabolism can be mitigated by making compounds more hydrophilic.

A problem ever-present is: do all of the compounds in a set react at the same site of action by the same mechanism? For example, many studies of the toxicity of phenols to all types of organisms (or their parts) have been made in the past 95 years. At present our database contains 127 QSAR for phenols; however, in these studies there was great bias toward lipophilic substituents with electron withdrawing character. These QSAR can be roughly lumped together as correlating "non-specific toxicity." However, it has recently been found that phenols exhibit a radical type of toxicity in rat embryos associated with strong electron realeasing substituents (σ^+) (*10*). A similar situation was found for anilines acting on fibroblast cells (*17*).

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Another example comes from a study of the mutagenicity of aromatic and heteroaromatic nitro compounds. Later it was possible to formulate a separate QSAR for them based on an established experimentally different mechanism which involved ring opening (*18*). These examples alert us to the fact that even rather simple compounds belonging to a single structural class may have quite different mechanisms of action depending on substituent selection. With good substituent selection subsets can be analyzed to establish homogenicity of mechanism.

Under-Parameterization and Equivocal Definition of Parameters

The simplistic parameterization in traditional QSAR is both a strength and a weakness. The QSAR can be rapidly formulated, if set design and substituent selection has been good, and readily interpreted. In fact many of those doing 3-D QSAR use a traditional QSAR for guidance (*18,19,20*). Overlooked in classical QSAR are special hydrogen bonding interactions, the dual nature of the hydrophobic effect (random walk and receptor interaction), metabolism and the complexity of steric interactions. Abraham, Taft and their colleagues are making real progress in formulating H-bonding parameters (*21,22*). It was recognized early on that metabolism (especially by P-450 enzymes) is often dependent on the hydrophobicity of chemicals so that to some degree logP and π terms account for it (*13*). Of course, this complicates the interpretation of the hydrophobic terms in the final equation.

How often do steric effects arise so that they can be recognized and accounted for by traditional QSAR? In the present database there are 218 QSAR containing E_S or E_S^C, 212 with MR and 89 with STERIMOL terms for a total of 519. In addition there are 287 with one or more indicator variables, many of which may be related to steric problems. While in the initial parameterization one attempts to include terms which, as experience suggests, may be useful in uncovering steric interactions, often it is only by trial and error that effective terms are brought to light. The same is true for electronic terms (σ, σ^-, σ^+, σ_I) and to a lesser degree for logP and π.

The weaknesses of classical QSAR, collinearity and under-parameterization are clear enough. What, if any, are its strengths?

Influence of Hydrophobicity on Bioactivity

The importance of hydrophobicity, even if ill defined, is widely appreciated. Indeed, classical QSAR has played an important role in developing this concept. Before the advent of QSAR in 1962 (*1*), the subject of hydrophobicity was largely neglected by those developing bioactive compounds or explaining their toxicity. It must be remembered that the operational definition of hydrophobicity from partition coefficients produces a complex parameter determined by the many ways in which water and octanol interact with organic chemicals. Studies by Taft, Abraham (*21,22*) and their colleagues are beginning to delineate some of these. Using four parameters (molecular volume, polarizability, H-bond accepting and H-bond donating), they can calculate logP very well if there are not too many hetero atom functions present. However, how the four parameters should be used to formulate QSAR nonlinear in π or logP is not yet evident. Clearly it would be most helpful if more hydrogen bonding parameters were available. It has been shown that a combination of logP and hydrogen bonding constants can be effective in the correlation of non-specific toxicity (*23*).

It has been tacitly assumed that, in the passive penetration of chemicals through hydrophobic membranes, a point is reached where it is the increase in hydrophobicity which causes a departure in linearity (2,13,24). The assumption is that if the compounds under consideration are relatively small (MW < 500) the break in linearity is not due to or enhanced by size (molar volume). This is clearly not true in a variety of examples of drugs acting on multiple drug resistant tumor cells (*25*). There may be other situations where compounds with molecular weight less than 500 display restricted access to their sites of action. With poor design, a set of congeners can be highly collinear in terms of molar volume (MR) and logP and this may obscure a role for size. Refining our understanding of the nonlinear dependence of activity on logP is urgent.

It is of interest to see how often activity is nonlinearly dependent on logP or π. Our database now contains 527 QSAR with logP or π based on the bilinear model and 270 based on the parabolic model (*24*). Together they constitute 27% of the examples in the database. About 38% of the whole organism (animals, plants and insects) QSAR contain nonlinear hydrophobic terms. One might expect QSAR from organelles and cells to show fewer such relationships; however, the percentage is the same, 38%. In the case of enzymic QSAR only 18% contain nonlinear hydrophobic terms. This may seem high since the random walk process is probably not often involved with pure enzymes. It may result from an enzyme containing large hydrophobic binding regions other than that of the active site. The primary reason for the nonlinear enzymic QSAR is that hydrophobic patches or pockets on enzymes are of limited size. This is normally reflected in a bilinear rather than a parabolic relationship.

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Advantages of Classical QSAR

Of course the ease of formation without the need of a very large computer means that QSAR can be quickly established for comparison with those obtained by other methods. The greatest advantage of traditional QSAR is that its use of mostly experimentally based parameters, which have been tested since the advent of the Hammett equation in 1935, greatly assist in interpreting the mechanistic basis of both chemical and biological reactions of series of compounds in the same or different systems (9,14,15,26). Until we can construct the subject of QSAR somewhat like that of mechanistic organic chemistry we are not constructing a science.

We find the vast majority (about 90%) of QSAR contain a logP or π term. Of those that do not contain such terms, 53% are enzymic QSAR. Beyond its presence we now have some idea of what the coefficient (h) with those terms should be. If we select all QSAR from our database with π or logP terms and then eliminate those with parabolic or bilinear terms, we find 1593 examples. Of these only 91 have coefficients greater that 1.2 (6%). If the cut-off is set at 1.4 there are only 51 examples (3%). The same situation holds when h is negative. Of a total of 190 examples where a negative h is associated with a logP or π, only 12 (6%) have slopes below -1.2. Also, outside the range of ±1.2, many of the slopes have large confidence limits. Hence in developing a new QSAR model one begins to think that something may be amiss if h is out of the range ±1.2. Such large coefficients may be modeling more than the usual hydrophobic effect, possibly a steric problem.

Information about expected values of h is particularly valuable in formulating QSAR with logP or π in bilinear form. Unless there is good spread in the dependent variable, one obtains unrealistic slopes using the bilinear model. Although the slopes are unreasonable, their optimum $\log P_O$ or π_O generally agrees with that from the parabolic model. Nevertheless, one must not be misled by improper slopes which are not uncommon to find using the bilinear model. However, the bilinear model is so valuable that we do not believe that it should be avoided because of the slope problem. Our knowledge of the usual limits in h provides valuable protection.

Under-parameterization may not be all bad. For instance, it has been shown in a number of enzymic QSAR, where the structure receptor has been established by X-ray crystallography, that a benzene ring of the ligands may flip 180° to place meta substituents in quite different environments. This was deduced from relatively simple QSAR in which hydrophobic and hydrophilic meta substituents behaved in an anomalous fashion. Molecular graphics examination of the structure of the active sites confirmed the deductions (*14*). One might well have expected the flipping to be the result of steric effects; in fact the root cause was the hydrophobicity or lack of it for meta substituents. It was easy to understand the problem because of the rather simple meanings attached to classical QSAR.

Beyond Classical QSAR

To go beyond the naive postulates of traditional QSAR, many researchers are focusing on solving the collinearity problem mentioned above. It can be said that fundamentally the properties we now call electronic, steric and hydrophobic all depend on electrons. We are simply being arbitrary by trying to factor them to fit our conventional thinking. It should be possible to separate the fundamental properties into principal components which can then be used for model building; easier said than done. There is a serious lack of a large enough database of property values of organic compounds to get the fundamental principle components. Even if we had ideal sets of components it would be a very long time, indeed, before we could reorient our thinking to use them to construct a science of QSAR. To get around the lack of experimental properties, one might attempt to formulate them from molecular orbital calculations. Here too we would have to resort to the use of PLS. Our feeling is that for quite some time to come we need to use parameters which relate to chemistry as we understand it. In this sense molecular orbital parameters are an important means for expanding the scope of traditional QSAR.

In the brave new world of 3-D QSAR and the oncoming QSAR based on pictures of receptors defined via virtual reality studies, most researchers agree that delineating the receptor binding pattern by superimpositioning of the compounds is the most difficult first step. Geometry optimization to find the best conformation is a time consuming operation. It is of interest to consider the Hammett equation in this respect. No consideration for conformation is used in its application other than that contained in benzoic acids, phenols or anilines and the t-cumyl chlorides used to define σ, σ^-, and σ^+. Nevertheless, the Hammett equation more than holds its own against present levels of molecular orbital calculations when the methods can be compared. Classical QSAR has followed this lead with surprising success. The search for adverse steric effects is only attempted after hydrophobic and electronic terms fail to do a reasonable job.

A complication, as in all QSAR, is to decide where to draw the line in structural variation so that all compounds in a given set are acting by the same biochemical mechanism. Cross validation can offer help on this problem.

Comparative Molecular Field Analysis

CoMFA is the most ambitious and popular aproach to 3-D QSAR (*27*). To avoid the obvious under-parameterization of traditional QSAR it formulates hundreds or even thousands of parameters from which one uses partial least squares (PLS) and regression analysis to select the minimum number of principal components (latent variables) to formulate QSAR. These are selected by an iterative process which generates the components by maximizing the degree of commonality between the structural parameter columns and the dependent variable. In the evaluation phase of a PLS iteration, selection of principle components is based on the improvement in its ability to predict.

One of the most difficult phases of a CoMFA analysis occurs in the calculation of the 3-D structures, selection of the conformations and their superposition for arrangement to a common binding pattern. This is very much a trial and error process (*28*) ending when the "best" arrangement has been found. An exhaustive search is not possible with today's computers, however powerful, particularly when the conformation of the congeners must be varied as well. The variables introduced in this process are reminiscent of the use of indicator variables in traditional QSAR. They are used because they improve the correlation and make it more self-consistent. The only real check on their validity is to make and test new congeners and see if they are well predicted.

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

The next step in CoMFA is to decide how to calculate the interaction energies between ligands and probes. The points for calculation are set by placing the molecules in a 3-D grid of points, the spacing of which may be critical. What the optimum distance between points (lattice spacing) should be is not yet clear, but it is clear that if spacing is too close or too far apart the final result will be impaired (*27*). A distance of 2 Å seems to be the most popular; however, recently it has been argued that 1 Å could be better (*28*). Halving the grid spacing greatly increases the computation time of a method already computationally time consuming. Again variables are established much like indicator variables and are largely "hidden" in the final result.

Another problem is to decide on a boundary at which contact point interaction calculations will be made and inside of which they will not be made (establishing a "parent" structure). Again, adjustable parameters enter the model and are hidden from view in the final model. Further problems arise in deciding just how large a value for negative interactions by large groups projecting beyond the shell surrounding the "parent" molecule can be accepted. Unreal steric effects are implied by 40 to 60 kcal interactions so that a practical cut off point needs to be assigned. Again arbitrary variables enter the model.

The choice of probes to model ligand-receptor interactions is still not settled. The default Tripos probes, a pseudo methyl group and a proton, are often used to model steric and electrostatic interactions. The proton probe would, to some degree, characterize hydrophobic and hydrophilic space. Weak interactions would be given by saturated carbon centers and strong interactions would be shown by heteroatoms and unsaturations. The group at Abbott laboratories has used Goodford's GRID program to add a third probe in the form of a water molecule to define hydrophobic interactions. Kim has recently provided convincing evidence for the importance of the water probe (*29*). Others are experimenting with adding logP as a separate variable. However, it has been found that it must be scaled by a factor of 10 to 1000 (the exact value is set by trial and error) or its role is overshadowed by the many latent variables. Actually, logP is not the best parameter for receptor interactions. It would only be appropriate when the ligands are being completely engulfed in a hydrophobic pocket or cleft. Receptor sites are often a mixture of hydrophilic and hydrophobic regions (*14*) so that π values are more suitable. One of the fascinating aspects of CoMFA is that high values of r^2 can be obtained with only two probes. This would imply that hydrophobic effects, which seem so important in traditional QSAR, are spurious. The lack of necessity for hydrophobic terms also occurs with Klopman's CASE and Enslein's TOPKAT programs. Collinearity is an insidious problem!

The calculation of two or three types of probe interactions at many grid points results in hundreds, even thousands, of potential variables for small sets of compounds (10 to 50). It is here that the PLS procedure to develope orthogonal latent variables becomes crucial. Most of the latent varables make such small contribution to the model that they can be easily eliminated. Deciding on a final smallest number can be done via r^2.

In contrast to classical QSAR where one struggles to find a few significant variables, CoMFA produces a surfeit, including those we have termed "hidden" or "forgotten" variables as well as the huge number of latent variables. What all in terms of chemical properties has been enfolded into each of these remains a mystery. At present, the numbers in a QSAR based on PLS are almost impossible to interpret. In the case of CoMFA, it is in the color 3-D contour maps that the treasure lies. These are sort of semi-quantitative regions showing favorable and unfavorable sites for what would seem to be electrostatic or steric interactions. It is difficult to estimate whether or not a proposed structural change would fit in any of these regions. Selecting new substituents or other structural changes is difficult since the contours do no relate to the parameters with which chemists normally work.

Cramer and others feel rather confident that CoMFA will not yield chance correlations. In fact, it seems likely that the outcome of a CoMFA analysis will be an

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

internally highly self-consistent result, somewhat reminiscent of the results from neural network analysis.

CoMFA relies heavily on cross validation and q^2, in which one acts as if the value of one or more data points is unknown. These are left out and then predicted by the new QSAR. This is similar to the jack-knifing process which is used to find outliers in classical QSAR. Of course, why a datapoint is not behaving like the others is never really clear. Is it misaligned? Does it operate by a different mechanism? Is the data in error or is there faulty parameterization? In CoMFA the final decision rests on cross validation. Actually, as Cramer points out (*27*), there is no magic number for q^2 and one needs "to develop an intuitive feeling for the significance of various q^2." The same holds true for r^2 in traditional QSAR.

The high self consistency one strives for with CoMFA may not be easy to attain, as a recent study by Cho and Tropska shows (*28*). They observed that q^2 is quite sensitive to the overall orientation of a set of structures and can vary by as much as 0.5 units when the orientation is systematically varied. This lack of stability appears to be the result of including many calculations from irrelevant grid points. To circumvent this problem, they used a lattice spacing of Å1, factored the grid points into 125 subsets and then ran CoMFA on each of these. Those subsets for which a satisfactory q^2 could not be obtained were discarded and a final CoMFA was then made with the stable sets.

The current method of interaction calculation overlooks the possibility of covalent bond formation which is assumed to be absent in most drug interactions. This may be more true for drugs than for agrochemicals and environmentally toxic substances. Some perspective can be obtained on this point by considering how many QSAR contain σ terms. A search of the bio database finds 627 QSAR with σ, 120 with σ^- and 75 with σ^+ terms for a total of 822 (27% of the database). The presence of such a term does not necessarily mean that covalent bond formation is occurring, but it is unlikely to occur without such a term showing up in the QSAR. Walsh pointed out many years ago (*30*) that over 100 enzymes had been shown to form covalent intermediates "suggesting there may be some general advantage to holding on to a portion of the substrate" in enzymatic processes. Although intended drug and pesticide interactions often do not involve covalent bond formation, for metabolic reactions it is essential. Also for mutagenesis, carcinogenesis, and electrophilic cancer drugs it is essential. In agrochemistry, inhibitors of enzymes such as cholinesterase often involve bond formations. In the future as more QSAR are developed to elucidate metabolic reaction mechanisms, a necessary step if we are to be more efficient in the design of less toxic chemicals, bond formation will become a point of central significance. At present, the CoMFA assumption that it can be neglected is a good assumption on which to base initial development. More complicated problems can be addressed later.

Another problem that has received little attention by those working with CoMFA is uncovering nonlinear dependence of activity, on especially π or logP, but also on other parameters. At present 880 QSAR (29%) in our database contain either a parabolic or bilinear term. It will be difficult to disentangle such relationships after the PLS step. This is not a minor point because in the design of bioactive compounds it is essential to know "when to quit" in increasing the hydrophobicity in a set of congeners. $\log P_O$ (optimum logP) is one of the most useful discoveries of classical QSAR.

Advantages of CoMFA

It is still too early to be sure of the advantages or the shortcomings of CoMFA. It is a highly complex evolving system of analysis which, as the work of Cho and Tropska suggests, can be improved. No doubt many improvements will be made in the calculation of the probe interactions. The new 3-D methodologies can benefit from the many classical QSAR studies as can new studies in the traditional mode. It is for this reason that we are working hard on the construction of a QSAR databank.

One of the greatest advantages of working with the new 3-D approaches is that they force us to think more about the horrendous complexity of the reactions of complex organic chemicals with receptors. This and the studies with molecular dynamics will gradually increase our understanding of how the basic reactions of biochemistry occur.

In a way, CoMFA is a more conservative approach to QSAR than traditional QSAR in that at present it does not attempt to cope with anything more complex than noncovalent interactions with a receptor. Traditional QSAR, in attempting to deal with cells and whole organisms, has a vastly more difficult task. One might view classical QSAR as a scouting methodology which, with little effort, finds the salient features of an SAR problem. Information gained in this way can guide the work in the more difficult 3-D approaches.

Most important, the new 3-D approaches promise better methods for dealing with conformational and steric effects, the weakest areas in traditional QSAR. For instance, Cramer (*27*) reanalyzed a set of anilides, carbamates and ureas inhibiting photosystem II in chloroplasts. The first QSAR was made by Mitsutake *et al.* (*31*). Cramer noted that the inclusion of a logP term in addition to the latent variables "generally" improved the correlation. A recent classical QSAR reanalysis (*32*) of Mitsutake *et al.* data showed that by dropping 5 of 74 data points a $(\log P)^2$ and a steric term could be omitted with little loss in correlation ($r^2 = .85$, $r^2 = 0.87$). For the set of 74, a sensible bilinear correlation could not be made; however, with the reduced set of 69 an equation linear in logP (with seven other linear terms) with a slope of 1.08 (±.16) was found. The slope compares with many other examples from different laboratories (*32*): 0.81 (±.24), 0.85 (±.25), 1.23 (±.38), 0.91 (±.21), 1.03 (±.19), 1.35 (±.23), 0.72 (±.11), 1.09 (±.30), 1.02 (±.06), 1.05 (±.20), 1.22 (±.18), 0.78 (±.18), 0.92 (±.27), 0.94 (±.22), 0.72 (±.23), 0.92 (±.42), 1.17 (±.18), 1.08 (±.22); mean = 0.99. In these 18 examples correlation was almost entirely on π or logP. These many lateral relationships assure one of the great importance of the hydrophobic interaction. Moreover the slope of 1 is to be expected if complete engulfment of the ligand in a hydrophobic receptor occurs (*14*). Crude as the logP parameter is, it continues to provide a very general kind of insight which is valuable in pesticide design (*33*) and drug design (*34-37*). Its use in the principle of "minimal hydrophobicity" in drug design provides guidelines in drug research before QSAR is even attempted (*38*).

No doubt CoMFA captures steric nuances missed by QSAR based on a few terms, but the QSAR readily relate data from different test systems in different laboratories on different compounds and tests in a readily understandable way. Ki Kim in his chapter has taken the first initiative to show that comparative QSAR may be possible using the results from CoMFA.

Those steeped in classical QSAR would like to avoid the use of the PLS procedure. It would be nice if the 3-D methods could be used to solve the steric and conformational problems. Then more sophisticated molecular orbital methods could be used to obtain electronic parameters which together with hydrophobic parameters could provide more easily understandable QSAR.

Conclusion

In the following chapters we see exciting new uses and advances in the rapidly expanding field of QSAR. Out of the various approaches will arise ever more powerful means for understanding how organic compounds affect the many forms of life as well as its innumerable components and from this basic knowledge the design of much better bioactive compounds will be possible.

A most important aspect of this field of analytical science, often overlooked, is developing skill in experimental design, both initially and as experimental results begin to accrue. If structural changes are made so that there is not significant variation in an electronic, steric or hydrophobic property, no program can find a role

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

for it. Also the analytical approach has greatly reduced redundancy in derivative preparation.

In closing, a few key references for agrochemical QSAR (*33,39-45*) and general QSAR (*5,23,34,38*) are cited.

Literature Cited

1. Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. *Nature* **1962**, 194, 178-180.
2. Hansch, C.; Fujita, T. *J. Am. Chem Soc.* **1964**, 86, 1616-1626.
3. Draber, W; Büchel, K. K.; Dickoré, K.; Terbst, A.; Pistorius, E. *Prog. Photosynthesis*, Metzner, H. Ed. Verlag C. Lichtenstein, Munich **1969**, III, 1789-1795.
4. Draber, W. *Z. Naturforsch.* **1987**, 426, 713.
5. Fujita, T. in *Comprehensive Medicinal Chemistry*, Ramsden, C. A. Ed., Vol. 4, Pergamon, **1990**, pp. 554-555.
6. Takayama, C.; Kamoshita, K.; Fujinami, A.; Kirino, O. in *Rational Approaches to Structure, Activity and Ecotoxicity of Agrochemicals*, Draber, W.; Fujita, T. Eds., CRC Press, **1992**, pp. 316-330.
7. Plummer, E. L. in *Pesticide Synthesis Through Rational Approaches*, Magee, P. S.; Kohn, G. K.; Menn, J. J. Eds., ACS Symposium Ser. 255, **1984**, 297-320.
8. See chapter by Chuman *et al.*
9. Hansch, C. *Acc. Chem. Res.* **1993**, 26, 147-153.
10. Hansch, C.; Telzer, B.R.; Zhang, L. *CRC Critical Rev. Toxicol.* **1995**, 25, 67-89.
11. Fujita, T.; Nishimura, K.; Cheng, Z. M.; Yoshioka, H.; Minamite, Y.; Katsuda, Y. in *Natural and Engineered Pest Management Agents*, Hedin, P. A.; Menn, J. J.; Hollingworth, R. M. Eds., ACS Symposium Ser. 551, **1994**, 396-406.
12. Fujita, T. in *Trends in QSAR and Modelling 92*, Wermuth, C. G. Ed., ESCOM, Leiden, **1993**, pp. 143-159.
13. Hansch, C. *Acc. Chem. Res.* **1969**, 2, 232-239.
14. Hansch, C.; Klein, T. *Acc. Chem. Res.* **1986**, 19, 392-400.
15. Debnath, A. K.; Shusterman, A. J.; de Compadre, R. R. L.; Hansch, C. *Mutat. Res.* **1994**, 305, 63-72.
16. Baker, B. R. *Design of Active-Site-Directed Irreversible Enzyme Inhibitors*, Wiley, **1967**.
17. Hansch, C. *6th International Workshop on Quantitative Structure-Activity Relationships in Environmental Sciences*, Belgirate, Italy, **1994**.
18. Debnath, A. K.; Hansch, C.; Kim, K. H.; Martin, Y. C. *J. Med. Chem.* **1993**, 36, 1007-1016.
19. Akamatsu, M; Nishimura, K; Osabe, H.; Uemo, T.; Fujita, T. *Pestic. Biochem. Physiol.* **1994**, 48, 15-30.
20. Osabe, H.; Morishima, Y.; Goto, Y.; Masamato, K.; Nakagawa, Y.; Fujita, T. *Pestic. Sci.* **1992**, 34, 27-36.
21. Abraham, M. H. *Chem. Society Rev.* **1993**, 22, 73-83.
22. Abraham, M. H.; Whiting, G. S.; Alarie, Y.; Morris, J. J.; Taylor, J. J.; Doherty, R. M.; Taft, R. W.; Nielson, G. D. *Quant. Struct.-Act. Relat.* **1990**, 9, 6-10.
23. Hansch, C.; Leo, A. *Exploring QSAR*, American Chemical Society, Washington, D. C., **1995**, p. 199.
24. Ref. 23, p. 188.
25. Selassie, C.D.; Hansch, C.; Khwaja, T.A. *J. Med. Chem.* **1990**, 33, 1914-1919.
26. Hoekman, D.; Leo, A.; Li, P.; Zhang, L.; Hansch, C. In *Proc. 27th Ann. Hawaii Int. Conf. on System Sciences*; Hunter, L., Ed.; IEEE Computer Society Press, **1994**, Vol. 5; p. 193-202.
27. Cramer III, R. D.; DePriest, S. A.; Patterson, D. E.; Hecht, P. in *3D QSAR in Drug Design*, Kubinyi, H. Ed., ESCOM, Leiden, **1993**, p. 443.
28. Cho, S. J.; Tropska, A. *J. Med. Chem.* **1995**, 38, 1060-1066.
29. Kim, K. H. *Quant. Struct.-Act.* **1995**, 14, 8-18.

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

30. Walsh, C. *Enzymatic Reactions Mechanisms*, Freeman, San Francisco, **1979**, p. 41.
31. Mitsutake, K.-I.; Iwamura, H.; Shimuzu, R.; Fujita, T. *J. Agric. Food Chem.* **1986**, 34, 725-732.
32. Ref. 23, pp. 462-473.
33. Clough, J. M.; Anthony, V. M.; de Fraine, P.J.; Fraser, E. M.; Godfrey, C. R. A.; Godwin, J. R.; Youle, D. in *Eighth International Congress of Pesticide Chemistry,* Ragsdale, N. N.; Kearney, P. C.; Plimmer Eds., Proc. American Chemical Society, Washington, D. C., **1995**.
34. Iwasaki, N.; Ohashi, T.; Musoh, K.; Nishino, H.; Kado, N.; Yasuda, S.; Kato, H.; Ito, Y. *J. Med. Chem.* **1995**, 38, 496-507.
35. Choe, Y. S.; Lidstrom, P. J.; Chi, D. Y.; Bonasera, T. A.; Welch, M. J.; Katzenellenbogen, J. A. *J. Med. Chem.* **1995**, 38, 816-825.
36. Benedini, F.; Bertolini, G.; Cereda, R.; Doná, G.; Gromo, G.; Levi, S.; Mizrahi, J.; Sala, A. *J. Med. Chem.* **1995**, 38, 130-136.
37. Altomare, C.; Cellomare, S,; Carotti, A.; Casini, G.; Ferappi, M. *J. Med. Chem.* 1995, 38, 170-179.
38. ter Laak, A. M.; Tsai, R. S.; den Kelder, G. M. D.-O.; Carrupt, P.-A.; Testa, B.; Timmerman, H. *Eur. J. Phar. Sci.* **1994**, 2, 373-384.
39. *Pesticide Synthesis Through Rational Approaches*; Magee, P. S.; Kohn, G. K.; Menn, J. J. Eds.; ACS Symp. Ser. 255, American Chemical Society; Washington, D. C., **1984**.
40. *Pesticide Chemistry: Human Welfare and the Environment*; Miyamato, J.; Kearney, P. C. Eds.; Pergamon Press; Oxford, **1983**, Vol. I, pp 339-376.
41. *Probing Bioactive Mechanisms*; Magee, P. S.; Henry, D. R.; Block, J. H. Eds.; ACS Symp. Ser. 413; American Chemical Society; Washington, D. C., **1989**.
42. *Safer Insecticides*; Hodgson, E.; Kuhr, R. J. Eds.; Marcel Dekker; New York, **1990**.
43. *Rational Approaches to Structure, Activity and Ecotoxicology of Agrochemicals*; Draber, W.; Fujita, T. Eds.; CRC Press; Boca Raton, FL, **1992**.
44. *Quantitative Drug Design*; Martin, Y. C.; Marcel Dekker; New York, **1978**.
45. *Drug Design: Fact or Fantasy?* Jolles, G.; Woolridge, K. R. H. Eds.; Academic Press; London, **1984**.

RECEIVED June 13, 1995

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch001

Chapter 2

Quantitative Structure–Activity Analysis and Database-Aided Bioisosteric Structural Transformation Procedure as Methodologies of Agrochemical Design

Toshio Fujita[1]

Department of Agricultural Chemistry, Kyoto University, Kyoto 606–01, Japan

We have been intensively studying the Hansch-type QSAR applications to a number of agrochemical series of compounds. Because the QSAR procedure utilizes principles of physical organic chemistry, clues for the molecular mechanism of action have been disclosed in many cases. From the QSAR models, new congeneric structures having the optimum activity profiles have been successfully predicted for some series. For generation of non-congeneric novel structures, we have been constructing a system named EMIL, which incorporates a database for structural "evolution" examples and a data-processing engine. In fact, the QSAR and EMIL procedures are complementary to each other under a category of computer-assisted empirical methodologies. In the QSAR procedure, the empirical model is built by mathematical equations describing correlations between variations in structure and bioactivity with use of physicochemical substituent and (sub)structural parameters. In the EMIL procedure, structural modification patterns, including those which are non-isometric, collected from the past structural evolution examples are used as empirical "rules" for "bioisosteric" structural transformations in a broader sense. The rules are applied to the primary lead structure to generate candidate structures having elaborated features. In this chapter, methodological backgrounds as well as characteristic distinctions of these procedures are presented on the basis of succesful topics for the agrochemical design.

The design of bioactive compounds including agrochemicals as well as pharmaceuticals could be defined as model-building with the use of information about and/or hypotheses based on structure-activity relationships and molecular mechanism-of-action studies. For synthetic chemists, the model should be attractive and motivational in helping them to increase the efficiency of syntheses of new candidate compounds whose bioactivity deserves to be explored toward further structural modifications and/or more advanced biological/pharmacological and field/clinical examinations.

---

[1] Current address: EMIL Project, Fujitsu Kansai Systems Laboratory, 2–2–6 Shiromi, Chuo-ku, Osaka 540, Japan

0097–6156/95/0606–0013$12.50/0

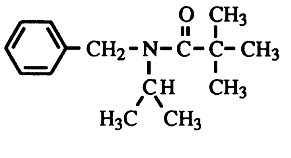
The classical QSAR (quantitative structure-activity relationship) procedure initiated by Hansch and his associates (*1,2*) provides us with the means to build the models based on concepts of physical organic chemistry (*3-5*). The model usually uses linear free-energy relationships (*6*) in a form of mathematical equations correlating variations in the observed potency of bioactivity with physicochemical molecular and substituent parameters among a series of congeneric compounds (*2*). The QSAR model has been successfully utilized in the design of bioactive compounds, some of them resulting in those currently on the market (*3,7,8*). A number of other QSAR examples are believed to exist in identifying candidate compounds brought into advanced developmental phases. Although many of them may be dropped for reasons such as undesirable side effects, environmental behavior, and selectivity spectra, this is beyond the scope of the QSAR methodology that eventually predicted the "best" candidates in earlier phases. Of course, toxicities and selectivities are often analyzable by other QSAR models if proper biological parameters can be selected (*3,4*). Sometimes, new candidate compounds can be designed and synthesized with "QSAR insights" extended outside the series of congeneric compounds with which the QSAR correlation equation is formulated (*9*).

It should be admitted, however, that the QSAR methodology is rather suitable for designing bioactive compounds within congeneric series. More or less drastic modifications beyond congeneric structures to obtain novel skeletal candidates are, if not by chance but at least with some design principles, mostly made with hypotheses based upon insights of practicing organic chemists into "qualitative" structure-activity patterns such as bioisosterism in a broader sense and information about biochemical modes of actions including the ligand-receptor interactions and the metabolic fates.

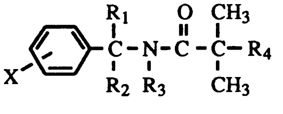
Recently, we have been developing a system named EMIL (Example-Mediated Innovation for Lead Evolution) to complement the above conditions of the QSAR methodology (*10,11*). The principal aim of the EMIL system is to present practicing chemists a computer-assisted procedure, which is not computational but entirely empirical, with the use of a database for "qualitative" but computer-readable information about "bioisosteric" modifications of bioactive compounds in a broader sense. In this article, we first present several topics including some "success" stories for agrochemical design as well as potential design principles for agrochemicals and related bioactive compounds. After comparison of the design principles used in these topics, we discuss the complementality of the QSAR and EMIL procedures.

Design of Agrochemicals and Related Bioactive Compounds

Topic 1 : Design of Bromobutide, a Selective Herbicide. Tebutam (**1**) has been used to control various broad-leaved and annual grass weeds for some 20 years (*12*). Following its launch from Gulf Oil Chemicals, a number of its analogs (**2**) have been synthesized and tested in various institutions. Among these, compounds **2**, in which X = H, $R_1 = R_2 = Me$, $R_3 = H$ and $R_4 = Me$ ~ *n*-Bu, are patented as being herbicidal against grass weeds (*13*), and compounds **2**, in which X = 3-Me or 2,5-Me_2, $R_1 = R_2 = R_3 = H$ and $R_4 = Me$, are possible candidates for paddy field herbicides (*14*).



1



2

One of the characteristic structural features of these compounds is that the amide function is shielded by bulky group(s) from either or both side(s).

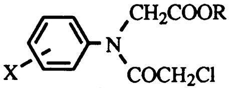
Kirino and coworkers of the Sumitomo Chemical noticed these previous results and tried to develop a new selective benzylamide herbicide to eradicate paddy field weeds. Before starting the project, the Sumitomo scientists also recognized the importance of steric bulk of substituents or substructures around probable functional moieties in some series of pesticides through their own studies. They were aware that the situations are analyzable quantitatively with the use of various types of steric parameters. For example, the fungicidal activity of the aminoacetonitriles (3) against a soil-borne disease of a Japanese radish species caused by *Fusarium oxysporum* increases with the bulkiness of R. For the 50% preventive dose, ED_{50} (mole), applied to a certain volume of soil where the radish is grown, equation 1 was formulated for 16 compounds where R is various normal, branched chain, and cyclo-alkyl groups (*15*).

$$pED_{50} = -\,0.61\ E_s^c + 1.52 \qquad n = 16,\ \ s = 0.20,\ \ r = 0.88,\ \ F(1, 14) = 50.1 \qquad (1)$$

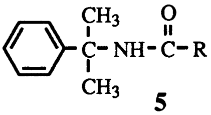
In equation 1, E_s^c is the Hancock "corrected" steric constant which is taken to be a parameter modeling the effect of α-branching of alkyl groups in addition to the steric bulk (*16,17*). The bulkier as well as the more α-branched the substituents, the more negative is the E_s^c value.

$RNHCH_2CN$

3



4



5

The herbicidal activity of the *N*-chloroacetyl-*N*-(substituted)phenylglycinates (**4**) against a paddy-field grass weed species, *Echinochloa crus-galli*, increases with the bulkiness of the *ortho* and *meta* X substituents on the benzene ring (*18*). For the 50% inhibitory concentration, I_{50} (M), against the shoot elongation, equation 2 was derived for the esters carrying various substituents at the *ortho* and/or *meta* position(s) (**4** : R = Et, X = *o*- and/or *m*-alkyl, -alkoxy, -halo, -acyl, $-NO_2$, $-CF_3$, and -Ph).

$$pI_{50} = -\,0.77\ E_s\,(ortho) - 0.22\ E_s\,(meta) + 3.99 \qquad (2)$$
$$n = 28, \quad s = 0.29, \quad r = 0.91, \quad F(2, 25) = 56.6$$

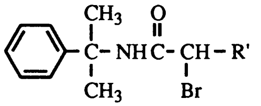
In equation 2, E_s is the Taft steric parameter (*16,19*). By definition, the bulkier the substituents, the more negative the E_s value. One of the phenylglycinate compounds, optimized in terms of the activity as well as ease of synthesis, is diethatyl ethyl (**4** : X = 2,6-Et_2, R = Et). Being introduced by Hercules Agrochemicals under the license of the Sumitomo Chemical, it had been practically used for the control of annual grass weeds, in particular, in broad-leaved crops such as sugar beets (*20*). The Sumitomo scientists considered that the steric bulk of substituents in the above series of compounds may play a role in protecting "pharmacophoric" sites from possible degradation mechanisms in soil and plants.

They transposed the above QSAR information about the "protective" steric bulkiness into the structural modification of benzylamide herbicides. They started the project from the α,α-dimethylbenzylamide structure (**5**) to increase the steric congestion/bulkiness of R systematically. The analysis for the first set of compounds, in which R varies from linear alkyls to such congested substituents as *t*-Bu, *neo*-Pent, CMe_2-*i*-Pr, CHBr-*t*-Bu, and CEt_2Me, afforded equation 3 (*21*).

$$pI_{50} = -\,0.15\ \pi^2 + 0.94\ \pi - 0.35\ E_s^c + 2.88 \qquad (3)$$
$$n = 41, \quad s = 0.27, \quad r = 0.93, \quad F(3, 37) = 83.0$$

In equation 3, I_{50} is the molar concentration required for 50 % inhibition against the shoot elongation of the seedling of the bulrush, *Scirpus juncoides*, a perenial weed species. π is the hydrophobic constant for aliphatic substituents (*22*). The compounds in which R = CH_2Cl and CH_2Br were not included in equation 3, because their activity profile differed from others including compounds in which halogens are not primary. Compounds **5**, in which R = CH(Hal)R', *i.e.*, halogens are secondary, were well incorporated in equation 3 as were compounds in which R is without halogens. The secondary halogen atom at the α-position in R does not seem to be labile enough to be attacked by nucleophiles and does not work to exert specific functions.

Equation 3 indicates that, although there is an optimum in hydrophobicity within the set of 41 compounds (π_{opt} = 3.3), there may be room for increasing the steric bulkiness of the R substituent in terms of (the negative) E_s^c to reach an optimum. Further structural modifications were made on the α-bromoacylamide structure **6**, because the compound **6** (R' = *t*-Bu) was observed to exhibit the highest potency in compounds included in equation 3.



6

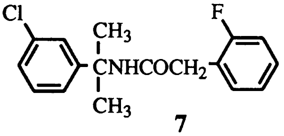
Too much congestion in R as tertiary substituents in structure **5** was not favorable to the activity. For 14 derivatives in which R' is varied from a simple alkyl to such highly congested/bulky groups as $-C(Me)_2$-*i*-Pr, $-C(Me)_2$-*i*-Bu and $-CMe(Et)_2$ in structure **6**, equation 4 was formulated to show that the optimum for the steric bulk of R' lies at about - 3.1 in terms of E_s^c (*21*).

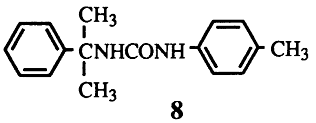
$$pI_{50} = -0.20\,(E_s^c)^2 - 1.23\,E_s^c + 4.39 \qquad (4)$$
$$n = 14, \quad s = 0.24, \quad r = 0.94, \quad F(2, 11) = 41.7$$

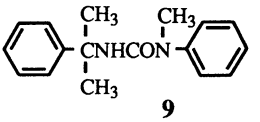
Equation 4 indicates that the entire R substituent in structure **5** should not be made extremely bulky. Beyond the optimum, the fit on its own receptor could be inhibited. As a matter of fact, the increase in the bulkiness/congestion beyond R' = *t*-Bu in structure **6** was observed to gradually decrease the herbicidal potency. Although the details are not shown here, quantitative analyses of the effect of substituents in the benzylamine moiety confirmed that the α,α-dimethyl substitution and the lack of the aromatic substitution are indeed favorable for the high activity against the bulrush shoot growth (*23*).

From the above analyses, bromobutide (**6** : R' = *t*-Bu) in which π(R) = 2.2 and E_s^c(R') = - 2.5, was selected as a promising novel paddy-field herbicide. Because of some collinearity between E_s^c and π values for compounds included in equation 3, the optimum π(R) does not coincide with that of bromobutide completely. The field trial showed that bromobutide is especially effective against the bulrush in paddy fields at rates of 100 - 500 g/ha. It has been commercialized since 1986 in Japan and southeastern Asian countries. Although it was included in the first set of compounds used to formulate equation 3, the decision-making for the selection of bromobutide would not have been made easily without the immediate follow-up of the bioassay results with the QSAR analyses during the course of series syntheses. The QSAR analyses in this example were by no means performed retrospectively after the synthetic project was over.

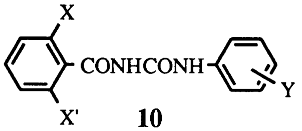
Kirino and his coworkers further modified structure **5** and found that such phenylacetamide derivatives as compound **7** are also effective against paddy field weeds (*24*). Structural features similar to those in bromobutide and compound **7** are found in daimuron (**8**) and methyldymron (**9**), commercial herbicides from Showa Denko in paddy fields (*25*) and turf protection (*26*), respectively.

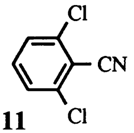

7

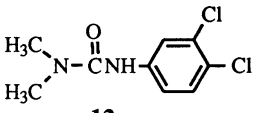

8


9

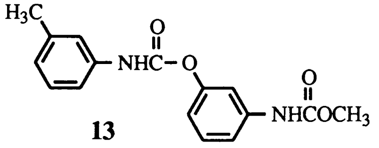
Topic 2 : QSAR and Optimization of Benzoylphenylurea Larvicides. Benzoylphenylurea derivatives are widely used larvicides to eradicate lepidopteran pest insects. Originally, the prototype compound (**10** : X = X' = Cl, Y = 3,4-Cl_2) was designed and synthesized (*27,28*) to obtain a novel herbicide by combining structural features of two types of herbicides, dichlobenil (**11**) and diuron (**12**) in Philip Duphar (now Solvey Duphar).


10


11


12

This type of strategy is not always effective. But, a precedent had been known leading to a successful product. Phenmedipham (**13**) introduced by Schering is the one in which the two herbicidal *N*-phenylcarbamate structures are combined (*29*).


13

Differing from the original expectation, however, compound **10** in which X = X' = Cl and Y = 3,4-Cl_2 was actually shown to have an insect growth regulating activity. An extensive QSAR examination was carried out by Verloop and Ferrell for the larvicidal activity against *Pieris brassicae* L. of a number of substituted compounds (*28*), showing that the 2,6-F_2 benzoyl substitution (**10** : X = X' = F) is better than the 2,6-Cl_2 (the modification to the 2,6-F_2-benzoyl compounds was initially made to reduce persistency in soil) and 4-substitution is more favorable to the activity than 3-substitution in the anilide moiety. The 4-substituent in the anilide moiety must be hydrophobic and electron-withdrawing as well as "thin" and "short". This QSAR examination has been taken to assist in the selection of diflubenzuron (**10** : X = X' = F, Y = 4-Cl) as the commercialized product.

We also synthesized a number of derivatives in which X, X' and Y in structure **10** are variously modified and measured the larvicidal activity against nondiapause larvae of the rice stem borer, *Chilo suppressalis* Walker, with a simultaneous application of piperonyl butoxide as the inhibitor of the oxidative metabolism. For mono- and di-*ortho*-substituted benzoyl derivatives in which X and X' are varied among halogens, lower alkyls and alkoxys, CF_3, NO_2, $N(CH_3)_2$ and CN, but Y is fixed as the 4-Cl, the di-*ortho*-substitution pattern is important not to allow an intramolecular hydrogen-bond formation for such *ortho* substituents as $N(CH_3)_2$ and OR with the NH hydrogen between two carbonyl groups on the double amide bridge (*30*). Electron-withdrawing and hydrophobic as well as small size substituents were indicated to be favorable leading to the 2-Cl and 2,6-F_2 substitutions that are among the best. The situation was formulated as shown in equation 5 (*30*).

$$\mathrm{pLD}_{50} = 0.77\ \Sigma E_s^{ortho} + 0.59\ \pi + 0.44\ \sigma^{\#} + 6.67 \qquad (5)$$
$$n = 20, \quad s = 0.26, \quad r = 0.90, \quad F(3, 16) = 21.4$$

In equation 5, ΣE_s^{ortho} is the sum of E_s values estimated experimentally from the rate constatnt of acidic hydrolysis of *ortho* substituted benzamides (*31*) over two *ortho* substituents, π is the difference in the experimentally measured log *P* value between (di)*ortho* substituted and unsubstituted benzamides (*32*), and $\sigma^{\#}$ is a composite electronic parameter defined so that $\sigma^{\#} = \Sigma\sigma_I$ for 2,6-disubstituted and $\sigma^{\#} = \sigma_{para}$ for mono-*ortho* substituted compounds (*32*). σ_I is the parameter for the inductive component of the electronic effect of substituents (*33*). LD_{50} is the dose (mmol/insect) required to kill 50% of larvae under certain experimental conditions.

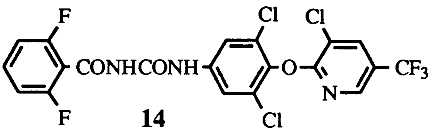
For derivatives in which X and X' are fixed as F and the 4-Y substituent is varied among halogens, CF_3, CN, lower alkyls and alkoxys, OPh, and substituted benzyloxys, equation 6 was formulated (*34*).

$$\mathrm{pLD}_{50} = 1.06\ \sigma_I + 1.57\ \pi - 0.17\ \pi^2 - 0.31\ \Delta B_5 + 4.93 \qquad (6)$$
$$n = 24, \quad s = 0.29, \quad r = 0.91, \quad F(4, 19) = 22.1$$

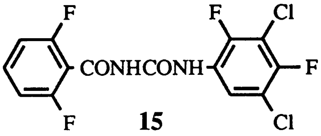
In equation 6, ΔB_5 is the STERIMOL parameter (*35*) which represents the maximum width from the axis connecting the α-atom of substituents with the rest of the molecule relative to that of H in the fully extended and staggered conformation, and π is the hydrophobic substituent constant taken from substituted acetanilide series (*34*). Equation 6 indicates that the higher the inductively electron-withdrawing property as well as the hydrophobicity (with an optimum at about 4.5 in terms of π) and the "slimmer" the shape of the Y substituent, the higher the activity (*34*), the result being slightly different from that of Verloop and Ferrell (*28*).

To optimize the structure according to the above requirements for the Y substituent simultaneously as far as possible, we designed and synthesized 4-(4'-substituted phenyl) and 4-phenylalkoxy derivatives (*34*). The higher phenylalkoxy substituents, which are hydrophobic and inductively electron-withdrawing, are not particularly relevant because the longer the alkyl chain, the chain-end benzene ring tends to deviate from the axis giving the larger ΔB_5 value. The 4'-bromophenyl substituent of which π is 2.72, σ_I is 0.15 and $\Delta B_5 = 2.11$, and the phenylpropoxy substituent of which π is 2.33, σ_I is 0.28 and $\Delta B_5 = 3.70$ showed a potent larvicidal activity, both being about 4 times as potent as diflubenzuron as far as the activity against the rice stem borer measured under conditions with piperonyl butoxide is concerned (*34*).

Recently, more complex benzoylphenylurea derivatives such as chlorfluazuron (**14**) from Ishihara (*36*) and teflubenzuron (**15**) from Celamerck (*37*) have been brought into practical uses. We showed that the larvicidal activity against the rice stem borer of these and many other derivatives with multiply substituted anilide moiety are quantitatively analyzable with an extended form of equation 6 using additional terms to represent position-specific steric and hydrophobic effects (*38*).



14



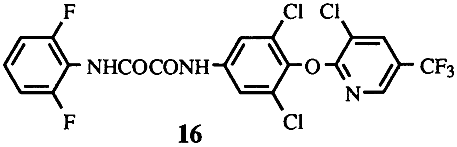
15

For 108 multiply substituted compounds covering those included in equation 6 and with various combinations of halogens, alkyls, alkoxys, substituted phenylalkoxys, substituted pyridyloxys, substituted phenyls, and CF_3, equation 7 was formulated (*38*).

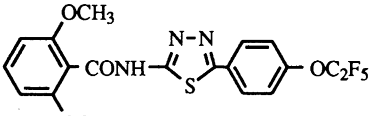
$$
\begin{aligned}
pLD_{50} = {} & 0.59\ \Sigma\sigma_I + 1.01\ \Sigma\pi - 0.12\ (\Sigma\pi)^2 - 1.27\ \pi(\textit{ortho}\text{-}2) - 0.68\ \Delta B_5\ (\textit{ortho}\text{-}2) \\
& + 0.27\ \Delta B_5(\textit{ortho})\cdot\Delta B_5(\textit{meta}) - 0.49\ \Sigma\Delta B_5(\textit{meta}\text{-}3,5) - 0.07\ \Delta B_5(\textit{para}) + 5.15 \qquad (7) \\
& n = 108, \quad s = 0.25, \quad r = 0.89, \quad F(8, 99) = 47.2
\end{aligned}
$$

The positions for position-specific effects of substituents are indicated in parentheses. The cross-product parameter, $\Delta B_5(\textit{ortho})\cdot\Delta B_5(\textit{meta})$, indicates a buttressing-type steric effect of *ortho* substituents on the vicinal *meta* substituents to make them trespass on the sterically least hindered *para* substituent region.

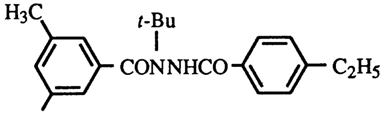
It is interesting to note that structures of experimental larvicides from Ciba-Geigy (**16**) (*39*) and Eli Lilly (now Dow Elanco) (**17**) (*40*) are considered to have emerged on the basis of structural features of benzoylphenylurea larvicides. Although the historical background of the discovery is its own, a recently marketed larvicide from Rohm and Haas, tebufenozide (**18**), shares the double amide bridge connecting two benzene rings with benzoylphenylureas (*41*).



16

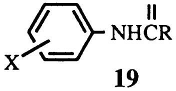


17

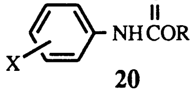


18

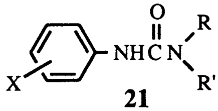
Topic 3 : Design of Cyclic Dicarboximide-type Fungicides. Compounds having the *N*-phenyl-amide moiety such as anilides (**19**), *N*-phenyl carbamates (**20**), and *N*-phenylureas (**21**) are herbicidally active exhibiting various degrees of the Hill reaction (a component of the photosynthetic system) inhibitory potency (*42, 43*). The most conventional substitution pattern on the benzene ring in these compound series is $X = 3,4\text{-}Cl_2$. Propanil (**19** : $X = 3,4\text{-}Cl_2$, R = Et), swep (**20** : $X = 3,4\text{-}Cl_2$, R = Me) and diuron (**21** : $X = 3,4\text{-}Cl_2$, R = R' = Me) are among representatives. They are regarded as being bioisosteric to each other.



19

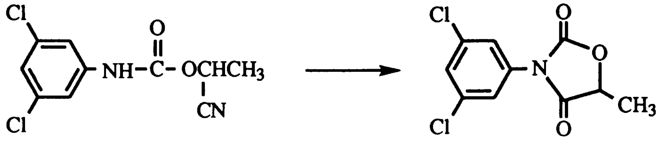


20



21

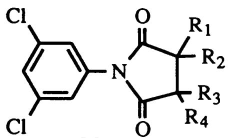
Fujinami and coworkers in the Sumitomo Chemical originally planned to make a novel herbicide of the type of structure **20**, in which $X = 3,5\text{-}Cl_2$ and $R = CH(CH_3)CN$, with utilization of lactonitrile ($CH_3CHOHCN$) obtained as a byproduct in manufacturing acrylonitrile (CH_2=CHCN) in the early sixties. The "carbamate" synthesized was found not to be herbicidal but showed a potent fungicidal activity (*44*). Because the fungicidal activity of the "carbamate" was not stable depending upon the lot of the "technical" samples, they soon realized that the active principle is not the "carbamate" itself, but something else transformed from the original carbamate during the storage. They isolated a compound the structure of which involves a cyclic dicarboximide moiety (**22**) being produced from the carbamate by a nucleophilic attack of the amide-NH lone-pair electrons on the cyano carbon atom followed by the hydrolysis (*44*).



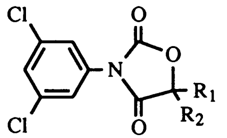
20 : X = 3,5-Cl_2,
R = $CH(CN)CH_3$

22

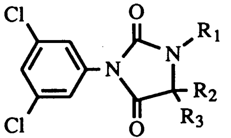
Following this discovery, a number of trials to simulate and optimize the structure **22** for improving the fungicidal characteristics have been made in a number of industries over the world (*44-49*). Syntheses and bioassays of series compounds **23** - **25** have led to launching at least three major commercial products, procymidone (**23** : $R_1 = R_4$ = Me, R_2 - R_3 = -CH_2-) from Sumitomo (*47*), vinclozoline (**24** : R_1 = Me, R_2 = CH=CH_2) from BASF (*48*) and iprodione (**25** : R_1 = $CONHCHMe_2$, $R_2 = R_3$ = H) from Rhone Poulenc (*49*), sharing the *N*-3,5-dichlorophenyl cyclic dicarboximide structure in common. They are particularly effective on *Sclerotinia* and *Botrytis* diseases in vineyards and greenhouses.



23



24

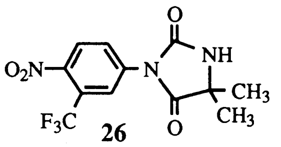


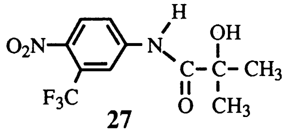
25

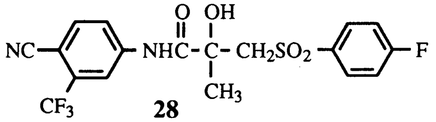
Structures of the cyclic imide moiety of above fungicidal compounds, the pyrrolidinedione (in **23**), oxazolidinedione (in **24**), and imidazolidinedione (in **25**), can be regarded as being generated through the cyclization of the side chain structures of the Hill reaction inhibiting anilides (**19**), carbamates (**20**), and ureas (**21**), respectively, with the insertion of another carbonyl component. Thus, structures **23** - **25** are bioisosteric. Regardless of the type of atoms next to the carbonyl function, the open chain bioisosteric "amides" (**19** - **21**) are the Hill reaction inhibiting herbicides and the ring-closured bioisosteric dicarboximides (**23** - **25**) are fungicides. In spite of the fact that the target of the biological activity and the optimum substitution pattern on the benzene ring differ from each other, the open chain "amides" and cyclic dicarboximides could also be regarded as being bioisosteric in a broader sense. Examples supporting this will be shown below.

Among anilides (**19**), chloranocryl (X = 3,4-Cl_2, R = -C(Me)=CH_2) and pentanochlor (X = 3-Cl,4-Me, R = $CH(Me)C_3H_7$) have been used practically to exterminate annual grass and broad-leaved weeds in various crop fields (*50*). They have the 3,4-disubstitution patterns as X as well as the branched chain alk(en)yl groups as R. Interestingly, a member of compound series **19** similar to the above herbicides, but having X = $3\text{-}CF_3,4\text{-}NO_2$ and R = $CH(Me)_2$ named flutamide from Schering, is an antiandrogen (*51*) and has been used as an anti-prostatic cancer agent for some 15 years. Flutamide, having the 3,4-disubstitution pattern on the benzene ring and the branched alkyl as R, is reasonably considered to show some Hill reaction inhibitory activity. Although no description about the herbicidal activity has been found, some higher homologs of flutamide in the acyl moiety have been observed to show a potent antibacterial activity (*52*).

Quite interestingly furthermore, compound **26** named niltamide from Roussel-UCLAF is also a potent and selective antiandrogen being used as an anti-prostatic cancer agent (*53*). The "bioisosteric" relationship between anilides and *N*-phenyl dicarboximides very similar to that described above in agrochemicals is observed in entirely different pharmacological category.

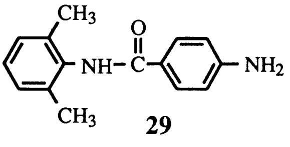

26

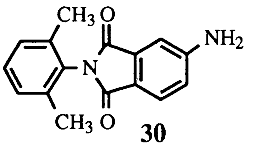

27

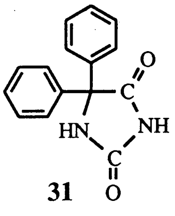

28

The dicarboximide heterocycle of nilutamide **26** belongs to the imidazolidinediones (in **25**). The structural differences of nilutamide **26** from the fungicidal compound series **25** are the substitution patterns on the benzene and imidazolidinedione rings. Flutamide works as its hydroxylated metabolite **27** *in vivo* (*54*). The hydroxy group in the metabolite **27** corresponding well with the NH group in nilutamide **26**; nilutamide is regarded also a ring-closured bioisoster of the metabolite **27**. By the way, bicalutamide (**28**) modified further from the "hydroxyflutamide" (**27**) is now being vigorously investigated for clinical use by Zeneca (*54*).

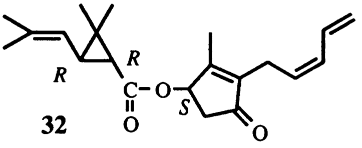
The above topic suggests that a "bioisosteric" structural transformation pattern in some series of compounds could be utilized as invaluable information about structural transformation in other series of compounds regardless of pharmacological categories. One of the recent examples of "bioisosterism" between amide and dicarboximide is that for the anticonvulsive benzanilide (**29**) and phthalimide (**30**) compounds (*55*). The activity is, respectively, comparable with and higher than that of phenytoin (**31**), a drug of choice for most types of epileptic disorders, in the anti-MES (maximum electroshock seizure) test in rats (*55*).


29

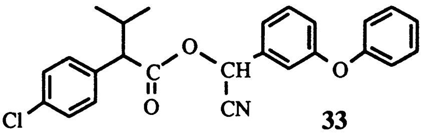

30


31

Topic 4 : Design of New Pyrethroid Series. Synthetic pyrethroids are a class of compounds the structure of which has been developed from that of insecticidally active principles of chrysanthemum flowers (*56*), the representative being pyrethrin-I (**32**). It is an ester, but the components are rather complex, the acidic component including a cyclopropane ring. Because of a lability being due to the cyclopropane structure under the agricultural weather conditions, the use of pyrethroids had mostly been limited to house-hold insecticides such as mosquito killers. A number of efforts have been made to modify the structure of acidic as well as the alcoholic component of pyrethrin-I (*56*). Fenvalerate (**33**), lacking the cyclopropane ring, launched by Sumitomo in 1983, has overcome this situation and expanded the utilization of synthetic pyrethroids to agricultural practice (*57*).

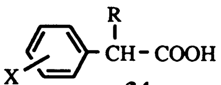


32

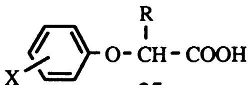


33

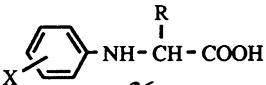
The acidic moiety of fenvalerate belongs to the α-alkylphenylacetic acids (**34**). It has been known that certain α-alkylphenylacetic acids exhibit an auxin activity along with their phenoxy (**35**) and phenylamino (**36**) analogs (*58*). A similar structure-activity relationship has also been recognized in DDT (**37**) and its analogs (**38** and **39**) (*59*). These examples indicate that substituted phenyl, phenoxy, and phenylamino groups are bioisosteric and interchangeable in certain instances.



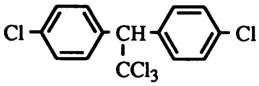
34



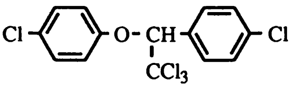
35



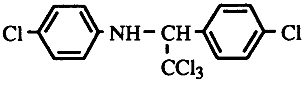
36



37

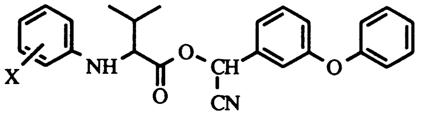


38

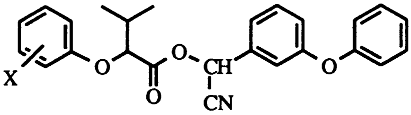


39

Our group including scientists from Dainippon Jochugiku applied this bioisosteric relationship to the structure of fenvalerate. The substituted phenylamino (**40**) as well as the substituted phenoxy analogs (**41**) of fenvalerate were synthesized and bioassayed showing that they are indeed active to various degrees against various pest insects (*60, 61*). An anilino analog (**40** : X = 2-Cl, 4-CF_3) named fluvalinate was also disclosed by scientists of Zoecon (*62*). As a matter of the fact, however, our syntheses and bioassays of "fluvalinate" and analogs had been made before the paper by the Zoecon group was publicized (*60*).

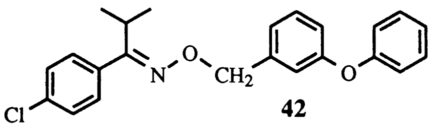


40

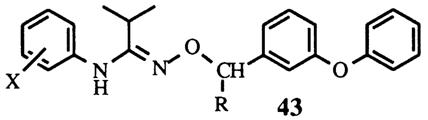


41

Being stimulated by the developments of fenvalerate and fluvalinate, further efforts to elaborate the structure of synthetic pyrethroids have been continued. We were interested in applying the structural modification pattern from fenvalerate to fluvalinate, *i.e.*, the bioisosteric relationship between substituted phenyl and phenylamino, to a non-ester type compound, SD47443 (**42**), which had been reported as being moderately active (*63*), to give an amidine compound (**43** : X = 2-Cl, 4-CF_3, R = CH_3) (*11*). The compound was not insecticidally active, unfortunately, but it showed a potent eradicating activity against house-dust mites and was patented (*64*). In compound **42**, the alcoholic component is not of the conventionally used α-cyanobenzyl alcohol-type. In this type of oxime-ether compound, a strong electron-withdrawing effect of the α-cyano group would make the molecule labile toward attacks of nucleophiles or basic biocomponents. In fact, no description of the α-cyano-benzyl analogs was found in the article reporting syntheses and bioassay results of SD47443 (**42**) and analogs (*63*). Because a similar lability was anticipated, we introduced the electron-donating methyl group as R in structure **43**.

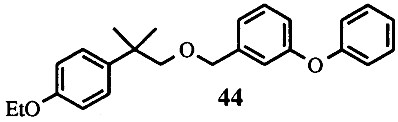


42

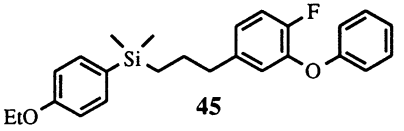


43

In the structural transformation from fenvalerate, efforts to modify the ester bonding are noteworthy, one of the examples of which is that leading to SD47443. The most recent examples include modifications to etofenprox (**44**) (*65*) and silafluofen (**45**) (*66*).



44



45

The above topics, covering diverse agrochemicals as well as design procedures, are believed to be neither specific nor biased so as to make generalized discussions possible.

Classification and Comparison of Design Procedures

The QSAR Procedure. Topic 1 describes an application of the QSAR procedure to the design of the herbicide, bromobutide (**6** : R' = *t*-Bu), which has been commercialized. It also includes the development of *N*-phenylglycine esters leading to the commercialized product, diethatyl ethyl (**4** : X = 2,6-Et_2, R = Et). The topic 2 is an example of the QSAR analyses covering a large number of congereric benzoylphenylurea compounds. In the course of analyses, we confirmed that the prediction by the QSAR for a smaller set of compounds is valid. With additional compounds including those practically used, the QSAR analyses were repeated and improved with a greater number of congeners leading to rationalization of complicated patterns of multiple substitution effects. The QSAR procedures in the topics 1 and 2 have illustrated how the bioactive potency is optimized for congeneric series of compounds by "systematically" introducing various substituents. This process of strucural transformation is usually called the lead optimization. The QSAR procedure is especially important and effective in the lead optimization phase in designs of bioactive compounds.

The representation of structure-activity relationship in the form of "linear" combination of physicochemical parameter terms in the QSAR procedure is indeed convenient in designing so as to make the activity more potent and optimal. Substituent modifications according to extraporation of certain parameter values toward directions enhancing the potency as well as to identification of a range of certain parameters optimizing the potency can be performed without difficulty. The physicochemical (substituent) parameters are either measured experimentally or calculated empirically with model systems.

The QSAR procedure can apply to such diverse systems as enzyme reactions, antimicrobial activities, pharmacological activities to excised tissues and organs, and the whole animal and plant bodies (*2-5*). It is almost impossible to clarify the microscopic as well as deterministic mechanisms responsible for the activity of bioactive compounds in each of the unit processes occurring in these diverse systems. As mentioned earlier, in the QSAR analysis, linear free-energy relationships between variations in structure and activity are examined. In other words, the procedure is regarded as being an extension of the Hammett-Taft-type analyses from physical-organic systems toward biological systems. The most useful advantage of the Hammett-Taft analyses is that, as one of the extrathermodynamic procedures (*6*), detailed identification of individual unit processes responsible for the (re)activity is not required in advance in the formulation of the correlation equations (*3*). Instead, from the correlation equations formulated, hypotheses for the molecular mechanism of (re)activities could sometimes be deduced empirically by comparative examinations with correlation equations for model (re)activities (*67*). A comprehensive compilation of physical-organic and biological QSAR examples has been made by Hansch and coworkers (*67, 68*).

The Procedure to Utilize Bioisosteric Relationships. The topics 3 and 4 do not include any QSAR principle. The core of each topic is to elucidate how the bioisosteric relationships are eventually "utilized" in designing new "skeletal" types of compounds. Examples are illustrated showing that the bioisosteric relationships are not necessarily limited within compounds exhibiting a particular biological activity but apply to a wider range of compounds of different pharmacologies. Structural-pattern variations in a set of bioisosteric structures observed in a certain series of agrochemicals are found in other series of agrochemicals as well as in pharmaceuticals.

To each of the topics, short introductory and a follow-up "stories" are attached. They seem to also include examples in which the bioisosteric structural transformations in the broader sense are eventually performed from the original or primary leads as well as from the commercialized "end" products. In these short examples and topics 3 and 4, it is believed that, the way of thinking of practicing organic chemists toward generating new lead structures is reflected. Their efforts are to make new skeletal structures maintaining at least the expected biological activity and hopefully exhibiting improved profiles from the initial lead. The structural transformation should not be too simple. However, too drastic transformation may be risky. Thus, the trial-and-error processes being inevitable, the successful structural transformations are likely to emerge empirically from accumulated experience and knowledge of as well as educated insights into structure-activity patterns and relationships of practicing chemists. If the trial is successful, a bioisosteric transformation is to be eventually made utilizing the bioisosteric relationship resulting in new lead structures.

To generate or identify new lead structures is perhaps the most challenging task for practicing chemists. However, the lead structures thus generated are often to be further modified to explore even newer lead structures. This type of structural modification is usually continued one after another sometimes in different places independently and/or competitively to eventually evolve "higher-ordered" or "new generation" structures from "primary" or "lower-ordered" lead structures. These consecutive structural modifications could be called the lead evolution (*9*).

Sometimes, the "rule" for the bioisosteric structural transformation is recognizable very easily as observed in structures of the Hill reaction inhibiting herbicides (**19** - **21**) and DDT analogs (**37** - **39**). However, each of the rules and examples seems to remain individually without being integrated. They have not been organized so that possible mutual relationships such as that found between Hill reaction inhibitors (**19** - **21**) and cyclic dicarboximide fungicides (**23** - **25**) may be overlooked. The original meaning of bioisosterism is believed to include the meaning of "isosterism" or "isometricity", *i.e.*, an equivalence in steric dimensions. The term, "bioisosterism," could be used as that in the broader sense. If the biological activity is maintained or even improved in compounds transformed from the initial leads, such transformations are to be called bioisosteric regardless of the difference in steric dimensions between initial and transformed compounds. The "equivalence" in the biological activity is taken to be more important than that in the steric dimensions. Thus, the "bioisosteric" transformations are made quite often without explicit perceptions for the bioisosterism in the lead evolution phase. In this respect, a number of "rules" for the non-isometric bioisosterism are still left to be integrated and systematized waiting to be explored in a number of successful structural transformation or lead evolution examples. Quite a few articles dealing with bioisosterism have been publicized (*4, 69-72*). They are, however, mostly collecting individual examples more or less restrained by the "isometricity", except for a few in which ideas to positively utilize the bioisosterism in the syntheses of novel compounds are presented (*4,72*). The system EMIL considers a way to integrate individual examples for the isometric as well as non-isometric bioisosteric structural transformation in the lead evolution phase expecting to promote the status as a systematized design procedure.

Collection of Bioisosteric Structural Transformation "Rules" and the Operation of the Transformation System

Identification of Bioisosteric Transformation "Rules". For the integration of bioisosteric structural transformation procedures in the broader sense, the EMIL system uses a database in which various lead evolution examples are accumulated in a computer-readable style (*10*). Each of the bioisosteric relationships in past examples is what has eventually been utilized as a "rule" for the bioisosteric transformation. In the system, the accumulated "example-based" rules are to be utilized to transform any lower-ordered lead structure. Thus, in each of the unit data in the EMIL database, the core of information is to identify the bioisosteric relationship between the lower-ordered and the higher-ordered structures in each example. Suppose compounds I and II are bioisosterically related, or the substructural modification of the compound I has eventually led to the compound II exhibiting the same type of bioactivity as compound I. The identification of substructural modification patterns is done by collating a substructure being modified in the structure I with a substructure having been modified in the structure II, leaving an unchanged substructural part or "evolutionary equivalent" counterparts between structures I and II (*10*).

Rules From Topic 1. In topic 1, the primary lead, tebutam (**1**), is "bioisosterically" modified to the secondary lead structure (**5** : R = *t*-Bu). The substructural modification pattern is indicated in Figure 1.

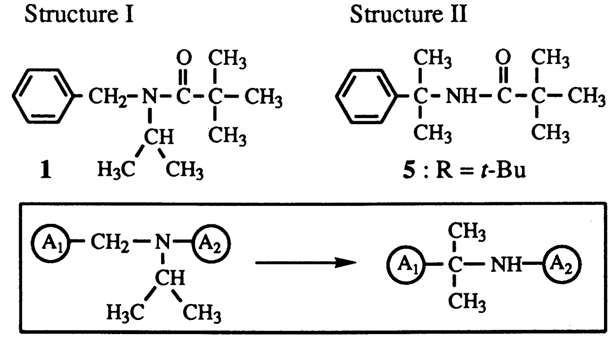


Figure 1. Substructural Modification Pattern in Bioisosteric Transformation of Tebutam.

In this example, the bioisosteric relationship as the rule is extracted so that the *N*-benzyl-*N*-*i*-propylamino structure is replaceable with the α,α-dimethylbenzylamino group. The benzyl-benzene ring being common to structures I and II, the rule is reduced to that enclosed in Figure 1 showing that the two bidentate substructures are interchangeable. Each of the circled A_1 and A_2 is unchanged or equivalent in structures I and II. The schematized rule as enclosed in Figure 1 is what is computer-readable.

Bromobutide (**6** : R' = *t*-Bu) has the structure in which the substructure A_2 in Figure 1 is optimized by QSAR. Further substructural transformations from bromobutide to compounds **7** - **9** could be schematized as shown in Figure 2.

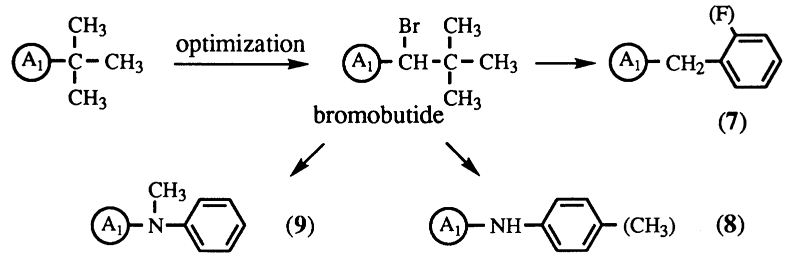


Figure 2. Substructural Modification Patterns in Bioisosteric Analogs of Bromobutide.

In Figure 2, A_1 is the α,α-dimethylbenzylaminocarbonyl which is common among analogs. In compound **7**, A_1 actually carries a *meta*-Cl substituent on the benzyl-benzene ring. The substituted phenyl groups within A_1 are regarded as being equivalent evolutionally with the unsubstituted parent. By the same token, the substituents on the benzene ring parenthesized in compounds **7** and **8** can be ignored. The selection of the best substituents on the benzene ring is supposed to be done in the optimization phase and the unsubstituted compounds should be regarded as new generation leads.

Rules From Topic 2. The initiation of topic 2 was the intuitive synthesis of the benzoylphenylurea (**10**) by combination of two herbicidal structures (**11** and **12**). The unexpected larvicidal activity of the benzoylphenylurea is entirely different from herbicidal activity of parent compounds. This type of structural modification could be called as being a sort of "bioisosteric" in the broadest sense, because the benzoylphenylureas exhibit important biological activity anyway. The situation can be described as shown in Figure 3.

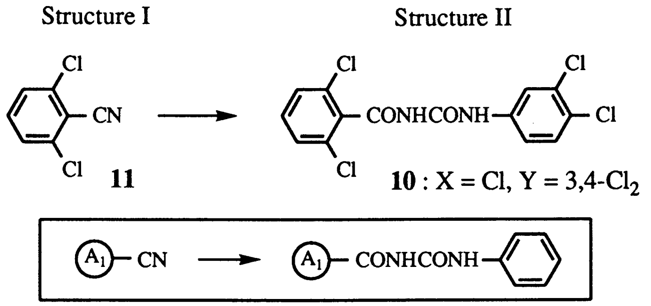

Structure I
Structure II
Cl
CN
Cl
11
Cl
CONHCONH
Cl
Cl
Cl
10 : X = Cl, Y = 3,4-Cl2
A1
CN
A1
CONHCONH

Figure 3. Substructural Modification Pattern from Dichlorobenil to Benzoylphenylurea.

The structural modification is not isometric at all. Moreover, the mode of biological action is entirely changed. However, this particular example is likely to be a precedent for the structural modification accompanied with an unexpected biological activity. The collection of this type of "rule" may be useful in cases when drasticities in structural modifications are permitted and new structures showing any biological activity are anticipated. Such "bioisosterism" with no dimensional equivalence as well as showing different modes of biological action could be called the "superbioisosterism".

The modification from diflubenzuron (**10** : X = F, Y = 4-Cl) to chlorfluazuron (**14**) is considered to be more "drastic" than the conventional substituent replacement. The "bioisosteric" relationship can be extracted as shown in Figure 4.

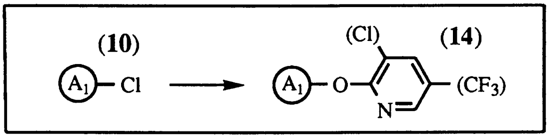

(10)
A1
Cl
(Cl)
(14)
A1
O
(CF3)
N

Figure 4. Modification Scheme from Diflubenzuron to Chlorfluazuron.

The structural variations from the benzoylphenylurea skeleton (**10**) leading to the oxamide (**16**), thiadiazole (**17**), and dibenzoylhydrazine (**18**) derivatives can be schematized as shown in Figure 5.

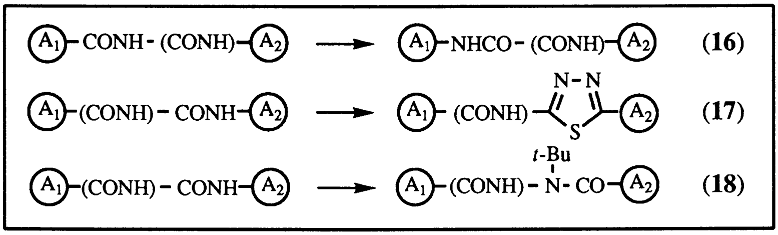

A1 - CONH - (CONH) - A2
A1 - NHCO - (CONH) - A2
(16)
A1 - (CONH) - CONH - A2
N-N
A1 - (CONH) - S - A2
(17)
A1 - (CONH) - CONH - A2
t-Bu
A1 - (CONH) - N - CO - A2
(18)

Figure 5. Substructural Modification Patterns from Benzoylphenylureas (**10**).

If the double-amide or amide-thiadiazole bridge is to be stressed as important substructures in the transformation rule, the parenthesized CONH structure can be saved. If not, it could be combined with the adjacent unchanged moiety to make new A_1 or A_2.

Rules From Topic 3. In topic 3, the herbicidal "amide" series of compounds **19**, **20**, and **21** are bioisosteric, as are fungicidal dicarboximide series of compounds **23**, **24**, and **25**. The situations can simply be schematized as shown in Figure 6.

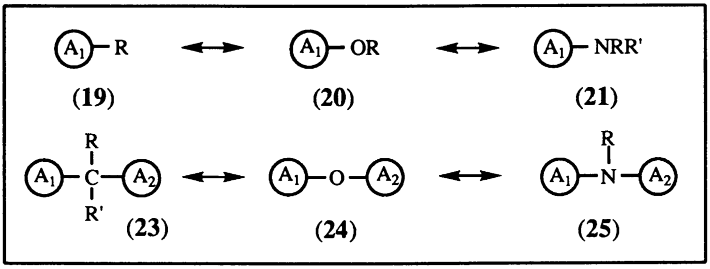


Figure 6. Bioisosterism among Alkyl, (Alk)oxy and Alkylamino Moieties.

As far as these two series are considered separately, the structural variations seem to follow isometric classical bioisosteric principles. The pyrrolidinedione (**23**) and imidazolidinedione (**25**) fungicidal structures are likely to be "designed" and synthesized on the basis of structures of anilide (**19**) and urea (**21**) herbicides, respectively, following the preceding example showing that the oxazolidinedione fungicides (**24**) are ring-closured analogs of the carbamate herbicides (**20**). The structural transformations common to these three cases are generalized as a single scheme shown in Figure 7.

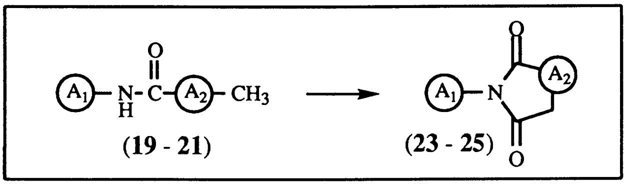


Figure 7. Structural Transformation from "Amides" to Dicarboximides.

Because biological modes of action differ between "amides" and cyclic dicarboximides, this could be one of the examples of the superbioisosterism. The "bioisosteric" relationships between flutamide [**19** : X = 3-CF_3,4-NO_2, R = $CH(Me)_2$] and niltamide (**26**) and between anticonvulsive compounds **29** and **30** seem to indicate, however, that the superbioisosteric relationship between herbicidal "amides" and fungicidal dicarboximides is not necessarily specific, but does not differ from the "conventional" bioisosteric relationships phenomenologically.

Rules from Topic 4. There are a number of evolutionary steps between pyrethrin-I (**32**) and fenvalerate (**33**). As far as the two compounds are concerned, the substructural modification patterns of acidic and alcoholic components are drawn as shown in Figure 8.

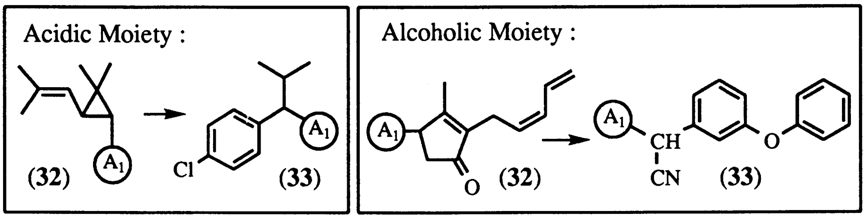


Figure 8. Substructural Modification Patterns from Pyrethrin-I to Fenvalerate.

Detailed transformation patterns which are more "isometric" can be explored in a number of evolutionary steps between two compounds.

The transformation pattern from fenvalerate (**33**) to fluvalinate (**40** : X = 2-Cl,4-CF_3) is formulated as shown in Figure 9.

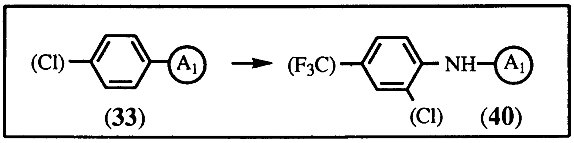


Figure 9. Transformation Pattern from Fenvalerate to Fluvalinate.

The pattern shown in Figure 9 is simply involved as one of the patterns for the bioisosteric relationship among phenylacetic (**34**), phenoxyacetic (**35**), and phenylaminoacetic (**36**) acids as well as among DDT (**37**) and its analogs (**38** and **39**), somewhat similar to that shown in Figure 6.

The pattern from fenvalerate (**33**) to SD47443 (**42**) can be schematized as shown in Figure 10.

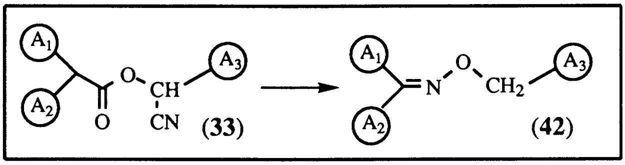


Figure 10. Structural Transformation Pattern from Fenvalerate to SD47443.

As a matter of fact, the second half of the topic 4 originates from the application of the EMIL system (*11*). The prototype structure (**43** : X = 2-Cl,4-CF_3, R = H) of the miticidal compound (**43** : X = 2-Cl,4-CF_3, R = Me) has been selected from a number of output structures generated by the EMIL system into which fenvalerate had been introduced as the input structure (*11*). A database for substructural modification patterns among synthetic pyrethroids was installed in the system and each pattern was applied to the input structure as the modification rule. Two rules shown in Figures 9 and 10 hit the structure of fenvalerate and that of the phenylamino analog, consecutively, affording the structure of the miticide prototype.

The transformations of fenvalerate (**33**) to etofenprox (**44**) and silafluofen (**45**) are shown with some simplification in Figure 11.

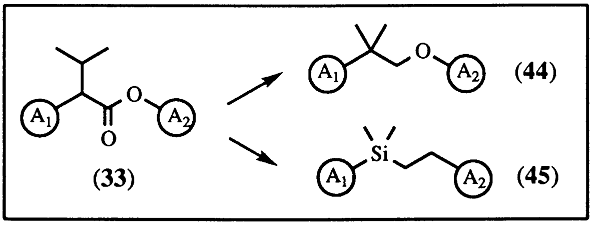


Figure 11. Structural Modification Patterns from Fenvalerate.

There are a number of the lead structure evolution examples other than those described above. They can be explored and collected so as to be utilized in the form of an integrated compilation such as the database. As described in some of the above examples, a certain bioisosteric relationship found for a set of compounds could sometimes apply to other sets of compounds regardless of the types of biological activity. Thus, the database had much better cover examples, not only from particular agrochemicals, but also others as well as pharmaceuticals. The rules found in the examples for pharmaceuticals could be utilized as the rules for the structural transformation of agrochemicals. The superbioisosteric relationships could sometimes be applicable without differentiation from the conventional bioisosteric relationships to explore "novel" compounds exhibiting bioactivity of any type. Even though we collected rules from existing successful topics retrospectively, the rules should be integrated and utilized prospectively for new trials.

Operation of the Bioisosteric Transformation System. The operational function of the EMIL system can be simplified as depicted in Figure 12 (*10,11*).

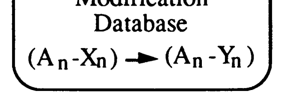

Primary "Lead" Structure
$[R_1-X_1]$
Input
Data Processing Engine
Output
Higher-ordered "Lead" Structure
$[R_1-Y_1]$
Substructural Modification Database
$(A_n-X_n) \rightarrow (A_n-Y_n)$

Figure 12. Simplified Operational Function of the EMIL System.

First, the structure of the primary lead compound, R_1-X_1, is introduced into the system. If an example, in which a structure A_1-X_1 is eventually transformed into A_1-Y_1, is hit by the database search, then, the system "automatically" constructs a candidate structure, R_1-Y_1, as that of the higher-ordered lead compound. The substructural modification pattern from X_1 to Y_1 originally identified in the structural evolution example from the structure I, [A_1-X_1], to the structure II, [A_1-Y_1], is utilized here as the rule for the substructural modification of R_1-X_1 to R_1-Y_1. Usually, more than a single patterns in the database are hit leading to a number of "brother" structures. The cycles of the operation can be repeated as far as the output structure R_1-Y_1 which is rewritten as R_2-X_2, is able to hit another rule with which A_2-X_2 is transformed to A_2-

Y_2 in the database. Of course, the symbol of structures does not mean that the "two" parts are monovalently combined. Instead, they are substructures in a certain structure.

Discussion

Although the output structures are constructed with substructural transformation rules extracted from existing lead evolution examples, the biological activity of compounds having these structures is not always guaranteed. One may also consider that most of the compounds with higher-ordered structures could be synthesized with various combinations of possible bioisosteric structures accumulated as the personal knowledge of expert practicing chemists without the aid of computerized data processing. Not every possibility could, however, be explored because of the limited memory of the human brain. Some promising candidate structures may be overlooked. The computer-assisted procedure is able to glean such structures. Moreover, the integration as a comprehensive compilation of the information about the bioisosteric structural transformations in the broader sense would be almost impossible without the aid of computer technology. Even though the system is only able to show a number of candidate structures for possible bioactive compounds, the bioactivity of which is not necessarily guaranteed, much more important is to gain insights into or hints as to how to elaborate further promising structures from the output structures.

As described above, in the process of designing bioactive compounds with structural modifications of the primary leads, there are at least two phases according to one's objectives : the one is lead generation or lead evolution to obtain novel skeletal compounds from which further elaborations are possible and the other is the lead optimization with systematic modifications of the lead structures. For the lead optimization phase, the QSAR procedure has been successfully employed as indicated in the topics 1 and 2. For the lead generation or evolution phase, the (super)bioisosteric relationships have been eventually utilized as described in the topics 1 - 4. The procedure to utilize the bioisosteric relationships has not been integrated in their original form. To integrate a number of (super)bioisosteric structural transformation, the EMIL system is trying to schematize and organize the individual information and to computerize the organized information as the database.

In the QSAR procedure, the prescription to optimize the lead structure is deduced from mathematical correlation equations. Therefore, it seems entirely different from the procedure used in the EMIL system. However, both of these procedures use empirical "rules". In the QSAR procedure, the rules are represented by variations in numerical parameters, while in the EMIL system, they are expressed by variations in (sub)structural patterns. Thus, within the category of computer-assisted empirical methodologies, the EMIL procedure could be complementary to the QSAR analysis. If the bioisosteric modification rules had been integrated and systematized so that the system such as the EMIL could be utilized ultimately, the entire story in the topics 1 - 4 could mostly be described in terms of the computer-assisted empirical methodologies.

The EMIL system can also be combined with such software systems as that to calculate the log *P* value and/or those to "predict" possible toxicities and environmental behaviors. Without sophisticated theoretical computations used in various computational procedures developed recently (*73*), this system could hopefully be well accepted smoothly by practicing chemists in such a way as the QSAR methodology has now been established (*4*).

Acknowledgments. The author wishes to thank the members of the EMIL working group who have devoted their effort to collect structural transformation examples and arrange them as the database. He also extends his gratitude to Dr. Yoshiaki Nakagawa of Kyoto University for his skilled assistance for the artwork and to Messrs. Noriyuki Shiobara and Toshihiko Kuboki of Fujitsu Kansai Systems Laboratory for their help for the schematization of some structural modification patterns.

Literature Cited

1) Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, F.; Streich, M. *J. Amer. Chem. Soc.* **1963**, *85*, 2817.
2) Hansch, C.; Fujita, T. *J. Amer. Chem. Soc.* **1964**, *86*, 1616.
3) Fujita, T. In *Comprehensive Medicinal Chemistry*; Hansch, C.; Sammes, P. G.; Taylor, J. B.; Ramsden, C. A., Eds.; Pergamon Press: Oxford, **1990**; Vol. 4, pp 497 - 560.
4) Hansch, C.; Leo, A. *Exploring QSAR*; American Chemical Society: Washington, D. C., **1995**; Vol. 1.
5) Kubinyi, H. *QSAR : Hansch Analysis and Related Approaches*; Methods and Principles in Medicinal Chemistry, Vol. 1; VCH Verlag: Weinheim, **1993**.
6) Leffler, J. E.; Grunwald, E. *Rates and Equilibria of Organic Reactions*; John Wiley: New York, NY, **1963**; pp 128 - 262.
7) Fujita, T. In *Drug Design* : *Fact or Fantacy?*; Jolles, G.; Wooldridge, K. R. H., Eds.; Academic Press: London, **1984**; pp 19 - 33.
8) Plummer, E. L. In *Pesticide Synthesis Through Rational Approaches*; Magee, P. S.; Kohn, G. K.; Menn, J. J., Eds.; ACS Symp. Ser. 255; American Chemical Society: Washington, D. C., **1984**; pp 297 - 320.
9) Fujita, T. In *QSAR in Design of Bioactive Compounds*; Kuchar, M., Ed.; J. R. Prous Science Publishers: Barcelona, **1992**; pp 3 - 22.
10) Fujita, T. In *Trends in QSAR and Molecular Modeling 92*; Wermuth, C. G., Ed.; ESCOM: Leiden, **1993**; pp 143 - 159.
11) Fujita, T.; Nishimura, K.; Cheng, Z.-M.; Yoshioka, H.; Minamite, Y.; Katsuda, Y. In *Natural and Engineered Pest Management Agents*; Hedin, P. A.; Menn, J. J.; Hollingworth, R. M., Eds.; ACS Symp. Ser. 551; American Chemical Society: Washington, D. C., **1992**; pp 396 - 406.
12) Schwartzbeck, R. A. *Proc. Br. Crop. Protect. Conf. -Weeds*, **1976**, Vol. 2, 739.
13) Buntin, G. A. *U.S. Patent*, **1970**-3498781 (*C.A.* **1970**, *72*, 111104t).
14) Shibayama, M.; Kishikawa, J.; Nakamura, H. *Japanese Unexamined Patent*, **1973**-88228 (*C.A.* **1974**, *80*, 104884j).
15) Kirino, O.; Oshita, H.; Oishi, T.; Kato, T. *Agric. Biol. Chem.* **1980**, *44*, 31.
16) Fujita, T.; Iwamura, H. In *Steric Effects in Drug Design*; Charton, M.; Motoc, I., Eds.; Topics in Current Chemistry 114; Springer-Verlag: Berlin, **1983**; pp 117 - 157.
17) Hancock, K.; Meyers, E. A.; Yager, B. J. *J. Amer. Chem. Soc.* **1961**, *83*, 4211.
18) Fujinami, A.; Satomi, T.; Mine, A.; Fujita, T. *Pestic. Biochem. Physiol.* **1976**, *6*, 287.
19) Taft, R. W., Jr. In *Steric Effects in Organic Chemistry*; Newman, M. S., Ed.; John Wiley: New York, NY, **1956**; pp 556 - 675.
20) *The Pesticide Manual, 9th Edition*; Worthing, C. R.; Hance, R. J., Eds.; British Crop Protection Council: London, **1991**; p 272.
21) Kirino, O.; Furuzawa, K.; Takayama, C.; Matsumoto, H.; Mine, A. *J. Pestic. Sci.* **1983**, *8*, 301.
22) Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley: New York, NY, **1979**.
23) Kirino, O.; Furuzawa, K.; Takayama, C.; Matsumoto, H.; Mine, A. *J. Pestic. Sci.* **1983**, *8*, 309.
24) Kirino, O.; Takayama, C.; Mine, A. *J. Pestic. Sci.* **1986**, *11*, 611.
25) Tashiro, M. *Jpn. Pestic. Inf.* **1979**, *No. 36*, 40.
26) Takematsu, T.; Konnai, M.; Ichizen, N. *Proc. 5th Asian-Pacific Weed Sci. Soc. Conf.* **1975**, 121.

27) Wellinga, K.; Mulder, R.; van Daalen, J. J. *J. Agric. Food Chem.* **1973**, *21*, 348, 993.
28) Verloop, A.; Ferrell, C. D. In *Pesticide Chemistry in the 20th Century*; Plimmer, J. R., Ed.; ACS Symp. Ser. 37; American Chemical Society: Washington, D. C., **1977**; pp 237 - 275.
29) Schulz, H. In *Progress in Photosynthesis Research*; Metzner, H., Ed.; Institute of Chemical Plant Physiology, Univ. of Tübingen: Tübingen, **1969**; Vol. 3, pp 1752 - 1760.
30) Nakagawa, Y.; Sotomatsu, T.; Irie, K.; Kitahara, K.; Iwamura, H.; Fujita, T. *Pestic. Biochem. Physiol.* **1987**, *27*, 143.
31) Sotomatsu, T.; Fujita, T. *J. Org. Chem.* **1989**, *54*, 4443.
32) Sotomatsu, T.; Nakagawa, Y.; Fujita, T. *Pestic. Biochem. Physiol.* **1987**, *27*, 156.
33) Charton, M. *Prog. Phys. Org. Chem.* **1981**, *13*, 119.
34) Nakagawa, Y.; Akagi, T.; Iwamura, H.; Fujita, T. *Pestic. Biochem. Physiol.* **1988**, *30*, 67.
35) Verloop, A. In *Pesticide Chemistry: Human Welfare and the Environment;* Miyamoto, J.; Kearney, P. C., Eds.; Pergamon Press: Oxford, **1983**; Vol. 1, pp 339 - 344.
36) Haga, T.; Toki, T.; Koyanagi, T.; Nishiyama, R. *J. Pestic. Sci.* **1985**, *10*, 217.
37) Becher, H.-M.; Becker, P.; Prokic-Immel, R.; Wirtz, W. *Proc. 10th Intern. Congr. Plant Protect. Abstract*, **1983**, 408.
38) Nakagawa, Y.; Izumi, K.; Oikawa, N.; Kurozumi, A.; Iwamura, H.; Fujita. T. *Pestic. Biochem. Physiol.* **1991**, *40*, 12.
39) Böger, M.; Drabek, J.; Neumann, R. *Brit. UK Patent Appl.* **1985**, GB2145716 (*C.A.* **1985**, *103*, 141636u).
40) Ward, J. S. *U.S. Patent*, **1979**-4141984 (*C.A.* **1979**, *90*, 186623s).
41) Hsu, A. C.-T. In *Synthesis and Chemistry of Agrochemicals II*; Baker, D. R.; Fenyes, J. G.; Moberg, W. K., Eds.; ACS Symp. Ser. 443; American Chemical Society: Washington, D. C., **1991**; pp 478 - 490.
42) Wessels, J. S. C.; van der Veen, R. *Biochim. Biophys. Acta* **1956**, *19*, 548.
43) Good, N. E. *Plant Physiol.* **1961**, *36*, 788.
44) Fujinami, A.; Ozaki, T.; Yamamoto, S. *Agric. Biol. Chem.* **1971**, *35*, 1707.
45) Fujinami, A.; Ozaki, T.; Nodera, K.; Tanaka, K. *Agric. Biol. Chem.* **1972**, *36*, 318.
46) Fujinami, A.; Tottori, N.; Kato, T.; Kaneda, N. *Agric. Biol. Chem.* **1972**, *36*, 1623.
47) Hisada, Y.; Kawase, Y.; Fujinami, A. *J. Pestic. Sci.* **1983**, *8*, 243.
48) Pommer, E.-H.; Mangold, D. *Meded. Fac. Landbouwwet. Rijksuniv. Gent,* **1975**, *40*, 713.
49) Lacroix, L.; Bic, G.; Burgaud, L.; Guillot, M.; Leblanc, R.; Riottot, R.; Sauli, M. *Phytiatr. Phytopharm.* **1974**, *23*, 165.
50) *The Pesticide Manual, 10th Edition*; Tomlin, C., Ed.; British Crop Protection Council: London, **1994**; p 789, 1066.
51) Sogani, P. C.; Whitmore, W. F. *J. Urol.* **1979**, *122*, 640.
52) Baker, J. W.; Bachman, G. L.; Schumacher, I.; Roman, D. P.; Tharp, A. L. *J. Med. Chem.* **1967**, *10*, 93.
53) Raynaud, J. P.; Azadian-Boulanger, G., Bonne, C.; Perronnet, J.; Sakiz, E. In *Androgens and Antiandrogens*; Martin, L.; Motta, M., Eds.; Raven Press: New York, NY, **1977**; pp 281 - 293.
54) Tucker, H.; Crook, J. W.; Chesterson, G. J. *J. Med. Chem.* **1988**, *31*, 954.
55) Bailleux, V.; Vallee, L.; Nuyts, J.-P.; Vamecq, J. *Chem. Pharm. Bull.* **1994**, *42*, 1817.

56) Yoshioka, H. In *Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals*; Draber, W.; Fujita, T., Eds.; CRC Press: Boca Raton, FL, **1992**; pp 185 - 217.

57) Yoshioka, H.; Ohno, N.; Hirano, M.; Fujimoto, K.; Hirai, H. *CHEMTECH* **1985**, *15*, 482.

58) Veldstra, H.; Booij, H. L. *Biochim. Biophys. Acta* **1949**, *3*, 278.

59) Hirwe, A. S.; Metcalf, R. L.; Kapoor, I. P. *J. Agric. Food Chem.* **1972**, *20*, 818.

60) Katsuda, Y.; Nakajima, M.; Fujita, T. *Japanese Unexamined Patent*, **1977**-82724; *Japanese Examined Patent*, **1982**-48522.

61) Katsuda, Y.; Nakajima, M. *Japanese Unexamined Patent*, **1977**-70021; *Japanese Examined Patent*, **1983**-2202.

62) Henrick, C. A.; Garcia, B. A.; Staal, G. B.; Cerf, D. C.; Anderson, R. J.; Gill, K.; Chinn, H. R.; Lavovitz, J. N.; Leippe, M. M.; Woo, S. L.; Carney, R. L.; Gordon, D. C.; Kohn, G. K. *Pestic. Sci.* **1980**, *11*, 224.

63) Bull, M. J.; Davies, J. H.; Searle, R. J. G.; Henry, A. C. *Pestic. Sci.* **1980**, *11*, 249.

64) Cheng, Z.-M.; Nishimura, K.; Fujita, T.; Minamite, Y.; Katsuda, Y.; Yoshioka, H. *Japanese Unexamined Patent*, **1993**-43536.

65) Utagawa, T.; Numata, S.; Oda, S.; Shiraishi, S.; Kodaka, K.; Nakatani, K. In *Recent Advances in the Chemistry of Insect Control*; Janes, N. F., Ed.; Royal Society of Chemistry: London, **1985**; pp 192 - 204.

66) Sieburth, S. M.; Manly, C. J.; Gammon, D. W. *Pestic. Sci.* **1990**, *18*, 289.

67) Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147.

68) Hansch, C.; Telzer, B. R.; Zhang, L. *CRC Critical Rev. Toxicol.* **1995**, *25*, 67.

69) Burger, A. *Prog. Drug Res.* **1991**, *37*, 287.

70) Lipinski, C. A. *Ann. Repts. Med. Chem.* **1986**, *21*, 283.

71) Hansch, C. *Intra-Sci. Chem. Rep.* **1974**, *8*, 17.

72) Thornber, C. W. *Chem. Soc. Rev.* **1979**, *8*, 563.

73) *Comprehensive Medicinal Chemistry*; Hansch, C.; Sammes, P. G.; Taylor, J. B.; Ramsden, C. A., Eds.; Pergamon Press: Oxford, **1990**; Vol. 4.

RECEIVED June 7, 1995

Chapter 3

Hydrophobicity Parameter of Heteroaromatic Compounds Derived from Various Partitioning Systems

Chisako Yamagami[1] and Toshio Fujita[2,3]

[1]Kobe Pharmaceutical University, Higashinada-ku, Kobe 658, Japan
[2]Department of Agricultural Chemistry, Kyoto University, Kyoto 606–01, Japan

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

The 1-octanol-water partition coefficient (log *P*), the chloroform-water partition coefficient (log P_{CL}), and the capacity factors (log k') in the reversed-phase liquid chromatography (RPLC) system of (un)substituted heteroaromatic compounds were measured. The relationships of log P_{CL} and log k' values with the log *P* value were analyzed in terms of hydrogen-bonding property of substituents. To find out the optimal conditions to estimate the log *P* value accurately from the RPLC system, the log k' values obtained with the ODS column using methanol-buffer (pH 7.4) mixtures as the mobile phase were examined as a function of the mobile phase composition. The use of eluents containing about 50% methanol gave the most straightforward correlation between log k' and log *P* except for compounds having amphiprotic substituents. In mobile phases in which the water content is higher, the relationship was complicated because of an increased difference in relative hydrogen-bonding effects of partitioning phases between the two systems. A new type of parameter S_{HA} for hydrogen-accepting substituents was proposed to correct their hydrogen-acceptability from solvents.

It is well known that the hydrophobicity is a very important factor affecting the biological activity. As a standard parameter of the hydrophobicity, the logarithm of the partition coefficient in the 1-octanol - water system, log *P*, is widely used (*1*), and a comprehensive compilation of log *P* values of various organic compounds is now available (*2*). However, not many studies have been performed concerning log *P* values of heterocyclic molecules (*3-5*) in spite of the fact that a number of biologically active compounds contain various types of heterocyclic rings. We have been accumulating log *P* values of (un)substituted heteroaromatic compounds and analyzing them in terms of substituent effects (*3,4*). Their partition behavior has been found considerably different from that of the benzenoid compounds mainly because the electronic interactions between the ring heteroatom(s) and the substituents cause variations in their hydrogen-bonding ability, and these are not always easily parameterized.

---

[3] Corresponding author. Current address: EMIL Project, Fujitsu Kansai Systems Laboratory, 2–2–6 Shiromi, Chuo-ku, Osaka 540, Japan

0097–6156/95/0606–0036$12.00/0

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Recently, log *P* values have also been estimated from capacity factors, log *k* ', measured with the reversed-phase liquid chromatography (RPLC) (*6-8*). Although this RPLC method is convenient and frequently used, standard conditions to estimate the log *P* value do not seem to be established. The accurate estimation of log *P* values of heteroaromatic compounds from the RPLC is expected to be difficult because hydrogen-bonding factors of the heteroatoms are sensitive to the variations in the mobile-phase solvent composition. To solve this problem, we thought that systematic analyses of log *k* ' values measured with various mobile-phase compositions would be necessary.

Previously (*9,10*), we have analyzed relationships among log *P* values of substituted benzenes, pyridines, and diazines measured in various partitioning systems. In this chapter, we first describe briefly the analysis of the relationship between log *P* and log P_{CL} for monosubstituted pyridines and pyrazines. Then, we extend our procedure for the analysis to the relationship between log *P* and log *k* ' values measured under various conditions for various series of heteroaromatic compounds. Finally, to obtain reliable log *P* value of heteroaromatic compounds having various substituents, a new hydrogen-accepting scale of substituents, S_{HA}, defined on the basis of the heat of formation under various dielectric environments, is proposed. Our emphasis is placed upon finding optimal RPLC conditions for estimating "accurate" log *P* values.

Methods

Partition Coefficients : The log *P* values for newly measured compounds were at 25°C by the conventional shake-flask method. Others were from literature (*2,3,5,11*).

RPLC Procedure : A Shimadzu LC9A liquid chromatograph equipped with SPD-6AV UV (Shimadzu) and SE-31 refractive index (Shoden) detectors was used. A commercial Capcell Pak C_{18} (4.6 mm x 15 cm, Shiseido) was used without further treatment. The MeOH-phosphate buffer (pH 7.4) eluents were prepared by volume. Retention times were measured using a C-R4A Chromatopac (Shimadzu). The capacity factor, *k* ', was determined from the retention time of each sample, t_R, and that of methanol, t_0, as $k' = (t_R - t_0) / t_0$. The log *k* ' value at 0% MeOH, log k_W, was calculated by the linear extrapolation from the plot of log *k* ' values against methanol concentrations ranging from 30 to 70% (*12*).

S_{HA} Parameters : A parameter, S_{HA}, representing the propensity of a certain substituted compound to be stabilized with being surrounded by dielectric media such as solvents relative to that of the unsubstituted compound was defined here for various substituents in various (hetero)aromatic systems. Thus, the S_{HA} value is a substituent constant specific to each of the skeletal systems. First, the minimum energy conformations of each compound in the gaseous state were calculated using the AM1 method (*13*) in the MOPAC 93 program package incorporated in an ANCHOR II modeling system (Fujitsu) (*14*). Then, using the minimum energy conformation in the gaseous state as the initial geometry, the conformational optimization and the calculation of the heat of formation in various solvents were made with use of the COSMO (conductor-like screening model) method (*15*) which approximates the effects of solvent molecules surrounding the molecule in question with the eps (ε: dielectric constant) command.

Thus, the heat of formation (H_f) values in the gaseous state (ε = 1), chloroform (4.8), octanol (10.3), methanol (32.7) and water (78.4) were obtained for each compound. When the H_f values calculated for five dielectric environments of each substituted compound (ArX) were plotted against the corresponding H_f values for the unsubstituted parent compound (ArH), a straight line was drawn as shown in Figure 1. The slope of this line was defined asS_{HA}..As observed, the higher the dielectric constant of the medium, the more negative the H_f value so that the greater the stabi-

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

lization induced by solvation with dipolar solvent molecules. As the X-substituents, either non-hydrogen bonding or hydrogen-accepting groups were used. Except for the gaseous condition as the reference state, the solvents examined were either hydrogen-donor or amphiprotic. Therefore, the S_{HA} parameter is expected to reflect the hydrogen-accepting propensity of the substituent X relative to H in a given series.

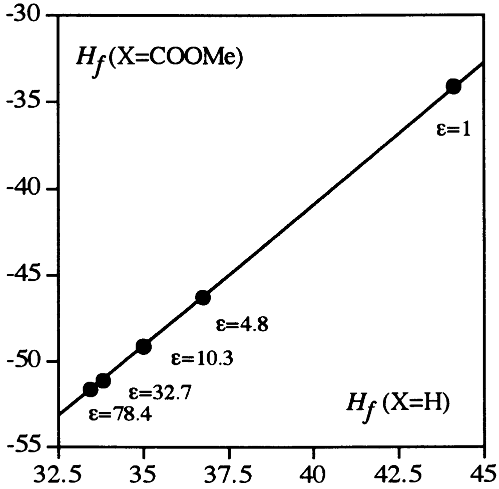


Figure 1. Plot of the H_f Values for a Substituted Pyrazine (X = COOMe) against Those for the Unsubstituted Pyrazine (X = H) Calculated under Various Dielectric Environments.

Results and Discussion

Relationship between log P and log P_{CL} : Previously, we have shown that the relationship between log P and log P_{CL} for disubstituted benzenes (X-C_6H_4-Y), where X and Y are variable and fixed substituents, respectively, is expressed in general as equation 1 (*9*).

$$\log P_{CL} = a \log P_{oct} + hHB + \rho\sigma^{o} + \text{const.} \qquad (1)$$

In equation 1, the parameter σ^{o} is the electronic substituent constant of the X-substituents free from the through-resonance effect. *HB* is an indicator variable which takes unity for H-accepting substituents X (for example, OR, Ac, CN, NMe_2, NO_2, and $CONMe_2$) and zero for non-hydrogen-bonding substituents (alkyls and halogens). Amphiprotic X-substituents are not included here because they usually show various degrees of the deviation from the regression line for non-hydrogen-bonders depending on the hydrogen-donating association of X-substituents with 1-octanol. The values a, h and ρ are regression coefficients. The value of "a" is close to unity. The $\rho\sigma^{o}$ term, when statistically significant, represents the electronic effect of X on the hydrogen-bonding solvation of the Y-functional group with $CHCl_3$ relative to that with 1-octanol. The *HB* term expresses the hydrogen-accepting effect of the hydrogen-accepting substituents at the substituent site. Usually, the h value is near 0.3. From detailed analyses (*9*), this value has been shown to correspond approximately to the logarithm of the ratio of the bulk solvent molality between $CHCl_3$ and octanol. The fact that the

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

discrete-type indicator variable, *HB*, works well suggests that the hydrogen-bonding equilibrium is almost entirely in favor of the bonded state for above *HB* substituents and the hydrogen-accepting equilibrium constant of these substituents in $CHCl_3$ is almost proportional or equivalent to that in octanol at least in conventional disubstituted benzene series.

Equation 1 has been applied to heteroaromatic compounds by regarding the ring hetero-atom(s) as the Y substituent(s) (*10*). Thus, the analysis for α-monosubstituted pyridines has produced equation 2.

$$\log P_{CL} = 1.04 \log P + 0.32HB + 0.42\sigma^{o} + 0.56 \qquad (2)$$
$$n = 15,\ r = 0.991,\ s = 0.092$$

In this equation and throughout this chapter, n is the number of compounds used for calculations, r is the correlation coefficient, and s is the standard deviation.

In the case of monosubstituted pyrazines, a correlation as shown by equation 3 has been formulated (*10*).

$$\log P_{CL} = 1.14 \log P + 0.37HB + 0.81 \qquad (3)$$
$$n = 15,\ r = 0.995,\ s = 0.067$$

In equations 2 and 3, the substituents used are alkyl, halogens, OR, SR, NMe_2, CN, Ac, CO_2R, and $CONMe_2$. Among them, the substituents, CO_2R and $CONMe_2$, were treated as substituents taking the *HB* value of two because they have two hydrogen-accepting sites. Moreover, in equation 3 but not in equation 2, weak hydrogen-acceptors such as OR, SR and NMe_2 were dealt with as if they are non-hydrogen-bonders ($HB = 0$), otherwise the correlation would be much poorer. This treatment is reasonable considering that the electron-withdrawing two ring *N*-atoms greatly decrease the hydrogen-accepting ability of such substituents in substituted pyrazines.

Relationship between log *P* and log *k'* : Most RPLC procedures have used a combination of the ODS stationary phase with the methanol-water mobile phase (*6,7*). From log *k'* values for a set of compounds, the non-measured log *P* value of a compound is estimated on the assumption that the log *P* value is linearly correlated to the log *k'*. This method is effective in cases when the solutes are congeneric and share a common hydrogen-bonding pattern.

One of the problems in dealing with the RPLC procedure is that the log *k'* value itself and the correlation with the log *P* value vary with the methanol content of the mobile phases. To eliminate the effect of the methanol composition, many investigators have conventionally used the log k_W value which is the extrapolated log *k'* value to the 0% methanol (*6,7*). In fact, this log k_W parameter is a direct indicator of log *P* for certain compounds which have no polar functional groups. However, we have found this approach ineffective in cases when hydrogen-bonding polar ring-hetero atoms and substituents are included in compounds.

First, we show the case of monosubstituted benzenes as the reference. In Figure 2, log k_W values are plotted against log *P*. Compounds with most non-hydrogen-bonding and "weakly" hydrogen-bonding substituents fit a single straight line of log k_W = log *P* meaning that, for these compounds, the log k_W value can be a direct indicator of the log *P*. However, carbonyl hydrogen-acceptors such as ester and acetyl substituents, gave a log k_W value about 0.3 higher than log *P*. In contrast, the amphiprotic phenol showed a lower log k_W value.

Next, heteroaromatic compounds in Table I were examined (*16*). They were categorized into three groups. The log k_W value for the group A compounds agreed well with log *P*. The log k_W value for the group B compounds was much larger than the log *P* value. On the other hand, for the group C compounds, the log k_W value was lower than the log *P* value. Group A compounds consist of aromatic rings with non-

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

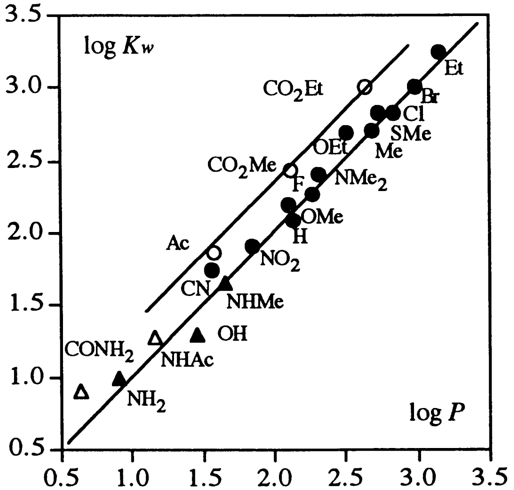


Figure 2. Relationship between log P and log k_W for Monosubstituted Benzenes. Closed circles : Non-hydrogen-bonders and "weak" hydrogen-acceptors, Open circles : Carbonyl hydrogen-acceptors, Closed triangles : Amphiprotics, and Open triangles : Amphiprotics having carbonyl.

hydrogen-bonding or very weakly hydrogen-accepting sites, group B compounds include heteroaromatic rings containing more than two hydrogen-accepting sites, and group C compounds contain a hydrogen donor. The above results for benzenoid and heteroaromatic compounds indicate that if we estimate log P value directly from the log k_W value on the assumption of the simple linear relationship, the value of hydrogen-acceptors tends to be overestimated and that of hydrogen-donors underestimated.

The important point, when using the RPLC method to estimate log P, is to find the optimal RPLC conditions that minimize the difference in the hydrogen-bonding factors from those involved in the 1-octanol/water system to produce a linear relationship between log P and log k' as far as possible. Keeping this in mind, we studied systematically the log k' value of (hetero)aromatic compounds.

When the relationship between log k' and log P of monosubstituted benzenes was examined with various mobile-phase compositions (at 15, 30, 50, 70 and 80% MeOH concentrations), the eluent containing 50% methanol, M50, gave the best linear relationship, where most substituents, except for amphiprotic substituents including OH, fell on the same straight line as shown in Figure 3. Although some amphiprotic solutes were failed to be lined up with others, we can conclude that M50 is the best eluent for estimating log P values for the monosubstituted benzene system.

The log k' values for heteroaromatic compounds and their alkylated derivatives included in Table I were plotted against log P in Figure 4. Interestingly, here again, the M50 eluent gives the best linear relationship, although group C hydrogen-donors deviate downward. In water-rich eluents, the group B strong hydrogen-acceptors deviate upward to form a second straight line parallel to that for the non-hydrogen bonders.

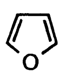
Similar hydrogen bonding effects were observed in correlations between log P and log k' for monosubstituted heteroaromatic series. In Figure 5, log k' values of monosubstituted pyrazines are plotted against the log P values (*17*). Also in this case, non-hydrogen bonders and hydrogen-acceptors are plotted as a single linear relationship

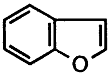
Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

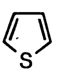
Table I. Log P and log k_W Values of Heteroaromatic Compounds[a]

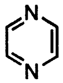
Compound	log k_W	log P	Δ^b	Compound	log k_W	log P	Δ^b
Group A							
PhH	2.01	2.13	-0.12	BF	2.70	2.67	0.03
Me-Ph	2.62	2.69	-0.07	1-Me-Pyr	1.17	1.15	0.02
Et-Ph	3.20	3.15	0.05	1-Me-In	2.65	2.64	0.01
FR	1.19	1.34	-0.15	TH	1.68	1.81	-0.13
2-Me-FR	1.80	1.85	-0.05	2-Me-TH	2.32	2.33	-0.01
2-Et-FR	2.42	2.40	0.11	3-Me-TH	2.29	2.34	-0.05
2,5-di-Me-FR	2.35	2.24	0.11	2-Et-TH	2.91	2.87	0.04
Group B							
PR	0.09	-0.26	0.35	PD	-0.33	-0.73	0.40
Me-PR	0.51	0.21	0.30	3-Me-PD	0.09	-0.35	0.44
Et-PR	1.00	0.69	0.31	4-Me-PD	0.22	-0.32	0.54
PM	-0.06	-0.44	0.38	QZ	1.22	0.90	0.32
2-Me-PM	0.29	-0.05	0.34	QX	1.51	1.32	0.19
4-Me-PM	0.38	-0.05	0.43	3-Me-TA	-0.24	-0.55	0.31
5-Me-PM	0.43	0.01	0.42	3-Ph-TA	2.03	1.71	0.32
Group C							
Pyr	0.55	0.75	-0.20	2-Me-In	2.35	2.53	-0.18
2-Et-Pyr	1.42	1.59	-0.17	3-Me-In	2.52	2.80	-0.28
2,5-di-Me-Pyr	1.33	1.47	-0.14	5-Me-In	2.49	2.68	-0.19
In	1.94	2.14	-0.20				

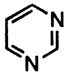
[a] Adapted from ref. *16*. Abbreviations are as follows:

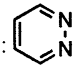
FR :


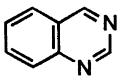
, BF :


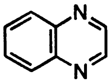
, TH :


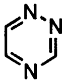
, PR :


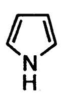
, PM :


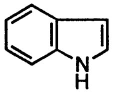
, PD :


QZ :


, QX :


, TA :


, Pyr :


, In :


[b] Difference between log k_W and log P.

with the M50 eluent. However, as the methanol concentration in the mobile phase is decreased, strong hydrogen-acceptors such as ester and *N,N*-dimethylamide groups deviate upward and the relationship becomes more complicated. The amphiprotic substituents deviate downward.

For monosubstituted furans and benzofurans, the log k' - log P plots are shown in Figure 6 (*18*). Here, we used alkyl groups, ester (CO_2R) and dimethyl amide ($CONMe_2$) groups, and amide groups (CONHR) as representatives of non-hydrogen-bonding, hydrogen-accepting, and amphiprotic substituents, respectively. In Figure 6, esters and dimethylamide together with non-hydrogen-bonders give a single linear relationship at M50, while the amphiprotic amide groups form a separate lower line with almost the same slope, the relationship being similar to Figure 5 for substituted pyrazines. The esters deviate upward with the decrease in the methanol content in the mobile phase.

We tried to analyze quantitatively the results described above and to derive correlation equations between log P and log k' at each eluent composition. Among various combinations of electronic constants and indicator variables to express the

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

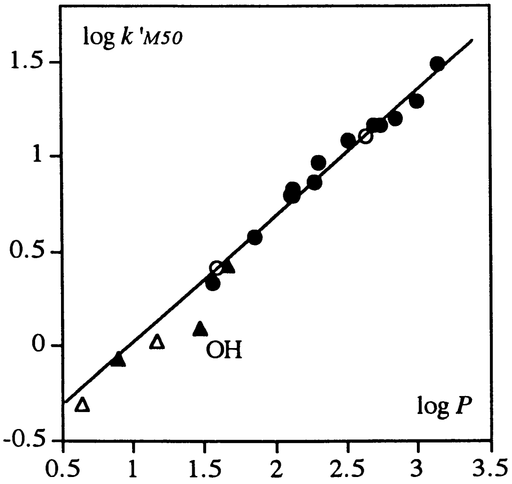


Figure 3. Relationship between log P and log k'_{M50} for Monosubstituted Benzenes. For symbols, see Figure 2.

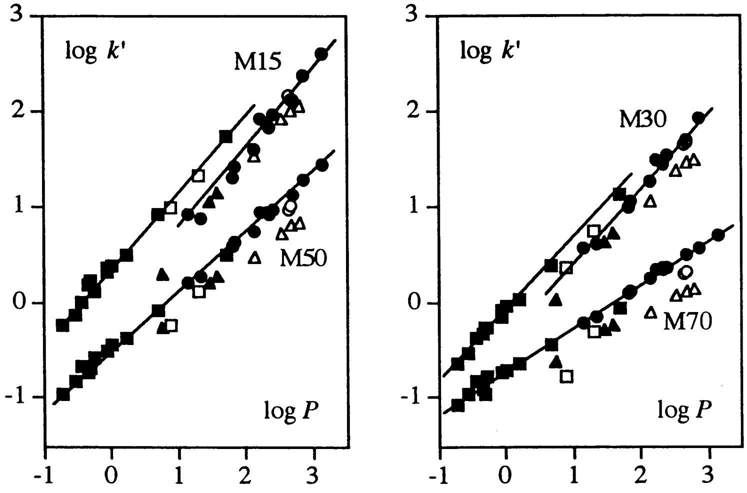


Figure 4. Relationship between log P and log k' Values for Heteroaromatic Compounds. Circles : Group A Compounds, Squares : Group B Compounds, and Triangles : Group C Compounds. Open symbols refer to benzo-derivatives. See Table I. (Adapted from ref. *16*.)

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

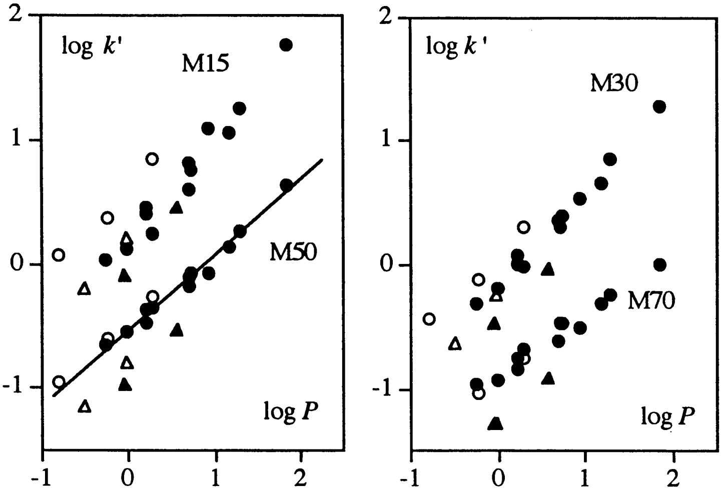


Figure 5. Relationship between log P and log k' for Monosubstituted Pyrazines. Symbols for substituents are the same as those in Figure 2. (Adapted from ref. *17.*)

hydrogen-bonding effects, we found that the log k' value of a series of compounds having variable substituents X can be described by the following general equation.

$$\log k' = a \log P + h_A HB_A + \rho\sigma_I + h_{AM} HB_{AM} + \text{const.} \qquad (4)$$

Here, σ_I represents inductive-electronic substituent constant of substituents (*19*). The HB_A and HB_{AM} are indicator variables for strongly hydrogen-accepting and amphiprotic X-substituents, respectively, which take the value of unity. The ester and *N*,*N*-dimethylamide substituents are classified into the HB_A category, while weak hydrogen-acceptors are classified as if they are either nonhydrogen-bonders or hydrogen-acceptors depending on the parent nucleus. The *HB* term functions as a correction for the hydrogen-bonding at the substituent site and the σ_I term works to represent the electronic effect of the substituents on the hydrogen-bonding of the ring hetero-atoms.

The correlations derived by equation 4 for monosubstituted pyrazines are given in Table II for the same data as those in Figure 5 (*17*). Very good correlations were obtained in every eluent composition. The contribution of the HB_A term is more important as the methanol concentration is decreased. The σ_I term seems to be needed only when the ring hetero atom is considerably hydrogen-bondable as in the case of pyrazine. The h_{AM} values are always negative, indicating that the amphiprotics are more "hydrophobic" in octanol-water partitioning system than in the chromatographic partitioning system. This is not unexpected if we consider that the amphiprotic substituents prefer to form the hydrogen-bond as a hydrogen-donor with the more "basic" octanol than the stationary phase.

The variations in the size of coefficient of each term with the eluent compositions in correlations obtained using equation 4 exhibited similar tendencies in other systems. Among them, the correlations for the M50 eluent are summarized in Table III. Irrespective of the skeletal structure of the unsubstituted compounds, the HB_A and σ_I terms were almost insignificant with the M50 eluent. Moreover, the coefficients "a" of the log P term are very close to each other, being 0.61 ± 0.04. This suggests that the contribution of the "hydrophobicity (in terms of the 1-octanol-water

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

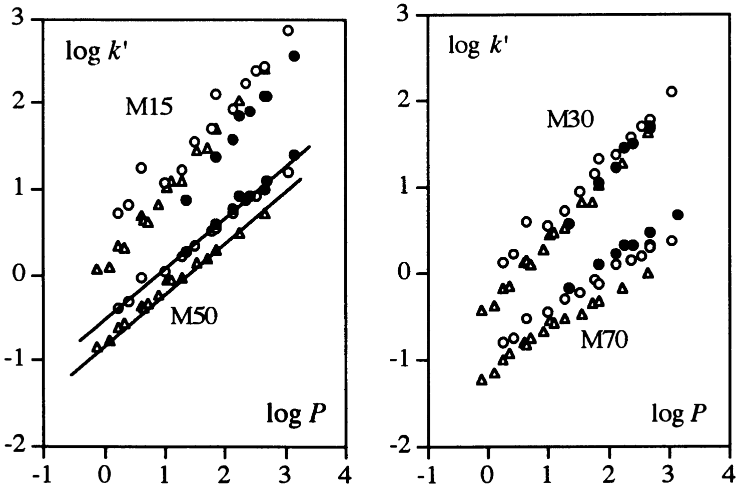


Figure 6. Relationship between log *P* and log *k*' for Monosubstituted-benzenes, -furans, and -benzofurans. Substituents are as follows. Closed circles : H and alkyl, Open circles : CO_2R and $CONMe_2$, Triangles : $CONH_2$ and CONHR.

Table II. Correlations for Monosubstituted Pyrazines according to Equation 4[a]

Eluent[b]	a	h_A	ρ	h_{AM}	const.	n	r	s
M15	0.824	0.414	-0.371	-0.319	0.292	19	0.991	0.078
M30	0.744	0.237		-0.305	-0.147	19	0.996	0.050
M50	0.582			-0.349	-0.511	19	0.992	0.060
M70	0.495			-0.348	-0.887	17	0.991	0.051

[a] Adapted from ref.*17*. Substituents included, Non-hydrogen-bonders : H, alkyls, F and Cl, "Weak" hydrogen-acceptors : OR, SMe, NMe_2, Ac, and CN, "Strong" or carbonyl hydrogen-acceptors : COOR and $CONMe_2$, and Amphiprotics : NH_2, NHMe, NHAc, and $CONH_2$. [b] The figure represents the MeOH % contained in the mobile phase.

system)" to the capacity factor can be separated from other factors by the simple treatment formulated as equation 4. Another point to be noted is that the h_{AM} value for heteroaromatic systems is more negative than that for the benzenoid systems (PhX and BC). This is probably because the electron-withdrawing property of the endocyclic hetero-atom(s) increases the hydrogen-donating ability of amphiprotic substituents in heteroaromatics. The above results demonstrate that log *k*' values obtained with the eluent containing 50% MeOH are usually correlated with log *P* by a single linear relationship as long as amphiprotic substituents are treated separately. It should be noted here that the behaviors of amphiprotics represented by a single indicator variable HB_{AM} in Tables II and III are because the variety of amphiprotic substituents is limited

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

Table III. Correlations Obtained by Equation 4 with the M50 Eluent

Series	a	h_A	h_{AM}	const.	n	r	s
PR[a]	0.582		-0.349	-0.511	19	0.992	0.060
5PM[b]	0.614	0.103[c]	-0.380	-0.423	14	0.999	0.024
FR[d]	0.635		-0.195	-0.544	18	0.999	0.022
TH[e]	0.620		-0.254	-0.644	27	0.996	0.055
PhX[f]	0.652		-0.118	-0.608	18	0.997	0.042
BC[g]	0.581		-0.067	-0.508	19	0.998	0.028

[a] Monosubst. pyrazine, for substituents included, see footnote a of Table II (ref. *17*).
[b] 5-Subst. pyrimidine, Non-hydrogen-bonders : H, alkyl, and halogen, "Weak" hydrogen-acceptors : OR, CN, and NO_2, "Strong" hydrogen-acceptors : CO_2R, and Amphiprotics : NH_2, NHAc, and $CONH_2$ (ref. *20*).
[c] HB_A was used only for CO_2R.
[d] 2- and 3- Subst. furans, Non-hydrogen-bonders : H and alkyl, "Weak" hydrogen-acceptors : OR and CN, "Strong" hydrogen-acceptors : CO_2R, and Amphiprotics : CONHR (ref. *21*).
[e] 2- and 3- Subst. thiophenes, Non-hydrogen-bonders : H, alkyl, and halogen, "Weak" hydrogen-acceptors : OR, CN, Ac, and NO_2, "Strong" hydrogen-acceptors : CO_2R, and Amphiprotics : CONHR.
[f] Monosubstituted benzenes, Non-hydrogen-bonders : H, alkyl, halogen, and CF_3, "Weak" hydrogen-acceptors : OR, SMe, CN, and NO_2, "Strong" hydrogen-acceptors or carbonyl hydrogen-acceptors : CO_2R and Ac, and Amphiprotics : NHAc and $CONH_2$.
[g] *m*- and *p*-substituted benzyl *N*,*N*-dimethylcarbamates, Non-hydrogen-bonders : H, alkyl, and halogen, "Weak" hydrogen-acceptors : OR, SMe, NMe_2, and NO_2, "Strong" hydrogen-acceptors : CO_2R, and Amphiprotics : CONHR, NHAc, and NH_2. (ref. *11*).

within CONHR, NHAc, and NHR (R : alkyl and H). With broader amphiprotic substituent species such as OH and COOH, more than a single indicator variable should be required (*9*).

Hydrogen-acceptor Scale : In above analyses, discrete-type indicator variables worked well for expressing hydrogen-bonding effects. As described above, however, the assignment of the value of zero or unity to the HB_A parameter is sometimes not explicitly preestablished, but dependent upon the (hetero)aromatic skeletal structure. Therefore, the utilization of the preestablished individual hydrogen-bonding parameter for each substituent was considered to be relevant. The S_{HA} value calculated for substituents X in various series of ArX (see Methods) was expected to be such a parameter. By definition, the S_{HA} value of H in any skeletal series is unity, and the higher the value for X, the higher the hydrogen-accepting propensity of the substituent X in a given skeletal series.

In Table IV, the S_{HA} value of fundamental substituents for monosubstituted pyrazines is given together with those for monosubstituted benzenes. Generally speaking, non-hydrogen-bonding and very weakly hydrogen-accepting substituents have the value near unity, and stronger hydrogen-accepting substituents have larger values. It is of interest to notice that substituents such as NMe_2 and OMe in the pyrazine series, which behave as non-hydrogen-bonders, have a S_{HA} value near unity, whereas the same substituents in the monosubstituted benzene series, which behave as hydrogen-acceptors, have a fairly large S_{HA} value.

We thought that the S_{HA} parameter could be a good indicator of hydrogen accepting ability as a first approximation. Thus, we re-analyzed our experimental data using this parameter. Regression analyses were made by replacing HB_A with S_{HA} and

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

Table IV. S_{HA} Parameter Value of Typical Substituents

Substituent	PhX	PR-X[a]
F	1.15	0.99
Cl	1.13	0.97
Me	1.03	0.96
OMe	1.92	1.02
SMe	2.13	0.95
CN	1.98	1.21
NMe_2	2.00	1.09
Ac	2.82	1.31
CO_2Me	2.96	1.64
$CONMe_2$	3.69	1.68

[a] Substituted pyrazines.

excluding the amphiprotic substituents from calculations. The results for monosubstituted pyrazines are given below. The relationship between log P and log P_{CL} shown by equation 5 was much improved by adding the S_{HA} term giving equation 6 as the counterpart of equation 3.

$$\log P_{CL} = 1.143 \log P + 0.855 \tag{5}$$
$$n = 15,\ r = 0.922,\ s = 0.254$$

$$\log P_{CL} = 1.113 \log P + 1.077 S_{HA} - 0.231 \tag{6}$$
$$n = 15,\ r = 0.993,\ s = 0.080$$

Similar analyses for the relationship between log P and log k' were performed using equation 7.

$$\log k' = a \log P + s S_{HA} + \rho\sigma_I + \text{const.} \tag{7}$$

The results given in Table V shows that excellent correlations can be obtained at every mobile-phase composition. The quality of correlations found in equation 6 and those in Table V is excellent, but almost equivalent to that of their counterparts using HB in equation 3 and Table II. In spite of the fact that the S_{HA} value is not available for the amphiprotic substituents, the use of the S_{HA} variable is advantageous because of no arbitrariness in this calculable parameter.

Table V. Correlations for Monosubstituted Pyrazines according to Equation 7

Eluent[a]	a	s	ρ	const.	n	r	s
M15	0.866	0.804	-0.472	-0.551	15	0.997	0.042
M30	0.769	0.448	-0.208	-0.575	15	0.998	0.034
M50	0.581			-0.511	15	0.990	0.059
M70	0.489			-0.884	14	0.986	0.052

[a] The figure represents the MeOH % contained in the mobile phase.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch003

The S_{HA} value calculated with the MNDO-PM3 hamiltonian was also tried, but the correlations were inferior in all the cases studied. Although semi-empirical calculations have limitations in their utility, the above results indicate that our new parameter is justifiable for expressing the hydrogen-accepting ability of substituents. Further investigation with a larger number of compounds will be needed to show its versatility. It should be noted that the S_{HA} parameter is ineffective in dealing with hydrogen donors or amphiprotics, and an effective method of treating hydrogen-donors is still to be found.

In this chapter, we showed how hydrophobicity parameters derived from different partitioning systems are correlated with their octanol-water log P values. Even if relationships seem complicated, they can be analyzed by using appropriate "correction" terms for hydrogen-bonding behaviors. Our analyses showed that in predicting log P values by RPLC, the use of eluents containing around 50% methanol seems to be more practical than the conventional log k_W approach. Especially in heteroaromatic systems, solutes with amphiprotic (H-donor) substituents yielded significant deviations from the regression lines for non-hydrogen bonders, and hence should be treated carefully.

Literature Cited.

1. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
2. Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR* ; American Chemical Society: Washington, D.C., 1995; Vol. 2.
3. Yamagami, C.; Takao, N.; Fujita, T. *Quant. Struct.-Act. Relat.* **1990**, *9*, 313 and the references cited therein.
4. Yamagami, C.; Takao, N.; Fujita, T. *J. Pharm. Sci.* **1991**, *80*, 772.
5. Bradshaw, J.; Taylor, P. J. *Quant. Struct.-Act. Relat.* **1989**, *8*, 279 and the references cited therein.
6. Braumann, T. *J. Chromatogr.* **1986**, *373*, 191.
7. Minick, D. J.; Frenz, J. H.; Patrick, M. A.; Brent, D.A. *J. Med. Chem.* **1988**, *31*, 1923
8. Terada, H. *Quant. Struct.-Act. Relat.* **1986**, *5*, 81.
9. Fujita, T.; Nishioka, T.; Nakajima, M. *J. Med. Chem.* **1977**, *20*, 1071.
10. Yamagami, C.; Takao, N.; Fujita, T. *J. Pharm. Sci.* **1993**, *82*, 155.
11. Yamagami, C.; Takao, N. *Chem. Pharm. Bull.* **1993**, *41*, 694.
12. Yamagami, C.; Takao, N. *Chem. Express* **1991**, *6*, 113.
13. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am.Chem. Soc.* **1985**, *107*, 3902.
14. Stewart, J. J. P.; Fujitsu Ltd., Tokyo.
15. Kamlet, A.; Shüürmann, G. *J. Chem. Soc., Perkin Trans. 2*, **1993**, 799.
16. Yamagami, C.; Takao, N. *Chem. Express* **1992**, *7*, 385.
17. Yamagami, C.; Ogura, T.; Takao, N. *J. Chromatogr.* **1990**, *514*, 123.
18. Yamagami, C.; Yokota, M.; Takao, N. *J. Chromatogr.* **1994**, *662*, 49.
19. Charton, M. *Prog. Phys. Org. Chem.* **1981**, *13*, 119.
20. Yamagami, C.; Yokota, M.; Takao, N. *Chem. Pharm. Bull.* **1994**, *42*, 907.
21. Yamagami, C.; Takao, N. *Chem. Pharm. Bull.* **1992**, *40*, 925.

RECEIVED June 27, 1995

Chapter 4

Theoretical Estimation of Octanol–Water Partition Coefficient for Organophosphorus Pesticides

Toshiyuki Katagi[1], Masakazu Miyakado, Chiyozo Takayama, and Shizuya Tanaka

Agricultural Chemicals Research Laboratory, Sumitomo Chemical Company, Ltd., 4–2–1 Takatsukasa, Takarazuka, Hyogo 665, Japan

A wide range of log*P* values for 67 organophosphorus pesticides possessing the various chemical structures were theoretically estimated by using parameters derived from their molecular geometries and electronic properties calculated with the MNDO-PM3 semiempirical SCF method. The multiple regression analysis showed that the log*P* values were satisfactorily expressed by van der Waals volume, total number of hydrogen bonding sites, and the LUMO energy. The equation correlated very well the log*P* values measured. Although the accuracy of the prediction is slightly lower than that of the CLOGP procedure, our method can estimate the log*P* values for the pesticides which could not be calculated by CLOGP because of the lack of the related fragment values.

The octanol/water partition coefficient (log*P*) of pesticides is one of the most basic physicochemical properties governing their biological activity as well as their environmental behavior in connection with the transport through various biological membranes and environmental phases (*1*). The most familiar technique for experimentally determining log*P* is the shake-flask method, and chromatography using the reverse-phase TLC or HPLC has been recently utilized as a convenient analytical method to avoid the intrinsic problem of the shake-flask method due to the formation of emulsions (*2*,*3*).

With the accumulation of experimental data, the fragment constant approaches represented by CLOGP have been successfully introduced (*4-6*) and their extensive applicability has made them a powerful tool estimating the log*P* value of a new compound. However, the following problems for complex molecules are now recognized: the way of dividing a structure into fragments, the lack of some fragment parameters, and the estimation caused by the ambiguous corrections for conformational and electronic effects. In order to solve these problems, various computational approaches have been undertaken by using quantum-chemical methods. One of them is based on the relation between log*P* and the molecular properties such

---

[1] Current address: Planning and Coordination Office, Sumitomo Chemical Company, Ltd., 27–1 Shinkawa 2-Chome, Chuo-ku, Tokyo 104, Japan

0097–6156/95/0606–0048$12.00/0

as a solvent-accessible surface area (*7-9*), electronic properties (*10-13*), and solvatochromic parameters (*14,15*). The others treat the partition as a thermodynamical process and directly calculate the corresponding free-energy change considering the solvation with the aid of molecular orbital (MO) methods (*16-20*). These approaches exhibit their usefulness in estimating the log$P$ values of small molecules. Incidentally, the specific programs developed by individual researchers are usually needed for these approaches and their applicability to complex molecules such as pesticides has not been studied extensively.

It is considered attractive to most chemists to obtain the log$P$ value of molecules from the descriptors easily obtained from chemical structures by using the popular MO programs. Based on these considerations, organophosphorus pesticides, most widely used in agriculture, possessing various chemical structures with a wide range of log$P$ were taken for our investigation as to whether simple MO calculations and descriptors familiar to chemists well afford theoretically log$P$ values.

Most organophosphorus pesticides dealt with here are the esters possessing the pentacovalent phosphorus and classified into the 17 classes from the atoms adjacent to phosphorus. Tables I-1 to I-6 summarize the classification of 67 compounds used in the analysis of the log$P$ values.

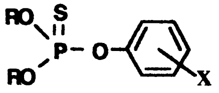
Calculations

The molecular geometry in lower-energy states together with the electronic properties was estimated by the MNDO-PM3 semiempirical SCF method (*21*). MNDO-PM3 is known to afford quite satisfactory results in the chemistry of the hypervalent phosphorus compounds (*22*). All calculations reported here were performed with the standard version of MNDO-PM3 in the MOPAC package (v. 6.0) of programs (QCPE 455). Geometries for stationary points were identified by minimization of the energy with respect to the geometric parameters using the BFGS algorithm (*23*) included in the MOPAC package. In order to calculate the solvent-accessible surface area (S_A) and volume (V_A), the program developed by Tomasi and his colleagues (*24*) (QCPE 554) was used for the PM3-optimized geometries. The radius of a sphere tracing out the surface was adjusted to 1.5Å as previously reported (*7,17*). Furthermore, the PM3-SM3 calculation developed by Cramer and Truhlar (*25*) in the AMSOL program (QCPE 606, v. 3.0.1) was used to estimate the molecular properties of hydrated pesticides. Unless otherwise noted, the experimentally measured log$P$ value was taken from the log$P$ file in the C-QSAR system 1.87 package (1994) developed by Hansch and Leo (*26*). The multiple regression analyses were done by using the QSAR program included in the ACACS system (*27*). In all analyses reported here, each term and equation are justified above 99.5% level by F and Student's t test.

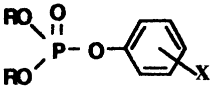
Molecular Geometries of Organophosphorus Pesticides

Although geometries and some electronic properties of small molecules are reported to be well reproduced by MNDO-PM3 calculation (*21*), its applicability to complex molecules such as pesticides is not clear. Therefore, we first examined whether the MNDO-PM3 is appropriate for our purpose by comparing the estimated geometries

Table I-1. Structure and LogP Value of Organophosphorus Compounds (Classes 1 and 2).



Class 1

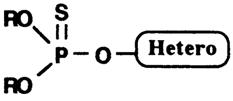


Class 2

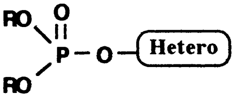
Class	#	Name	R	X	Obsd.[a]	Calc.[b]	CLOGP
(1)	**1**	Fenitrothion	CH_3	3-CH_3-4-NO_2	3.30	3.380	3.209
	2	Tolclofos-methyl	CH_3	2,6-Cl_2-4-CH_3	4.56	4.220	4.854
	3	Cyanophos	CH_3	4-CN	2.75	2.734	2.505
	4	Methyl Parathion	CH_3	4-NO_2	2.86	2.949	2.790
	5	Bromophos	CH_3	4-Br-2,5-Cl_2	5.21	4.557	4.951
	6	Chlorthion	CH_3	3-Cl-4-NO_2	3.45	3.398	3.345
	7	Fenchlorphos	CH_3	2,4,5-Cl_3	5.07	4.553	4.961
	8	Iodofenphos	CH_3	2,5-Cl_2-4-I	5.51	4.924	5.491
	9	Fenthion	CH_3	3-CH_3-4-SCH_3	4.09	4.066	3.907
	10	Dicapthon	CH_3	2-Cl-4-NO_2	3.58	3.469	3.545
	11	DEPPS[c]	C_2H_5	H	3.46	3.471	3.425
	12	Parathion	C_2H_5	4-NO_2	3.83	3.762	3.468
	13	Dichlofenthion	C_2H_5	2,4-Cl_2	5.14	4.636	5.033
	14	Fensulfothion	C_2H_5	4-$SOCH_3$	2.23	2.965	2.244
(2)	**15**	DMPP[c]	CH_3	H	1.22	1.223	1.337
	16	Oxon of (**1**)	CH_3	3-CH_3-4-NO_2	1.69	1.798	1.799
	17	Oxon of (**2**)	CH_3	2,6-Cl_2-4-CH_3	2.66[d]	2.785	3.444
	18	Oxon of (**3**)	CH_3	4-CN	0.84[d]	1.360	1.095
	19	Oxon of (**4**)	CH_3	4-NO_2	1.33	1.553	1.380
	20	Oxon of (**6**)	CH_3	3-Cl-4-NO_2	1.83	2.234	1.935
	21	DEPP[c]	C_2H_5	H	1.64	1.805	2.015
	22	Oxon of (**12**)	C_2H_5	4-NO_2	1.98	2.328	2.058
	23	Propaphos	*n*-C_3H_7	4-SCH_3	3.67	3.457	3.354

[a]Representative log*P* value in the CLOGP file. [b]Calculated by Eq. 5. [c]DEPP: *O,O*- diethyl phosphate, DMPP:*O,O*-dimethyl phosphate, DEPPS:*O,O*-diethyl phosphorothioate. [d]Determined by the shake-flask method in our laboratory.

Table I-2. Structure and LogP Value of Organophosphorus Compounds (Classes 3 and 4).



Class 3

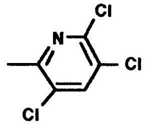


Class 4

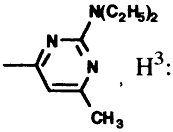
Class	#	Name	R	Hetero	Obsd.[a]	Calc.[b]	CLOGP
(3)	**24**	Chloropyrifos-methyl	CH_3	H^{1c}	4.31	3.762	3.764
	25	Pirimiphos-methyl	CH_3	H^{2c}	4.20	3.383	3.199
	26	Chlorpyrifos	C_2H_5	H^{1c}	5.27	4.362	4.442
	27	Diazinon	C_2H_5	H^{3c}	3.81	3.565	3.502
	28	Triazophos	C_2H_5	H^{4c}	3.55	4.054	3.114
(4)	**29**	Oxon of (**27**)	C_2H_5	H^{3c}	2.07	2.023	2.092

[a,b]See footnotes of Table I-1. [c]Structures are shown below.

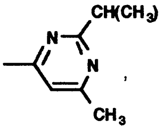
H^1:



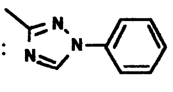
H^2:



H^3:

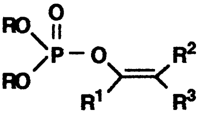


H^4:



with those determined by X-ray crystallography from the Cambridge structural database (*28*) with some additional compounds the crystallographic structure of which is known. Compounds used for these comparisons are shown in Table II. The initial structure of pesticides was generated from the typical geometries by using ACACS. The most extended form was taken for each moiety of the molecule. The optical isomerism around phosphorus and carbon atoms, if present, was defined to be identical with that observed in the crystallographic structure. To seek out the stable conformers, the conformational analysis was first conducted for each set of the two

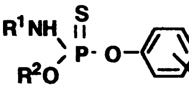
Table I-3. Structure and LogP Value of Organophosphorus Compounds (Class 5).



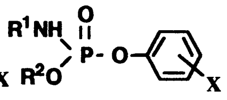
Class	#	Name	R	R^1	R^2	R^3	Obsd.[a]	Calc.[b]	CLOGP
(5)	**30**	Stirofos	CH_3	2,4,5-Cl_3-Ph	Cl	H	4.31	3.762	3.764
	31	Dicrotophos	CH_3	CH_3	H	$C(=O)N(CH_3)_2$	0.00	0.246	0.215
	32	Crotoxyphos	CH_3	CH_3	H	$COOCH(CH_3)Ph$	3.30	3.171	3.057
	33	Dichlovos	CH_3	H	Cl	Cl	1.43	1.002	1.809

[a,b]See footnotes of Table I-1.

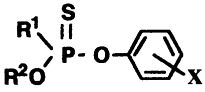
Table I-4. Structure and LogP Value of Organophosphorus Compounds (Classes 6 - 9).



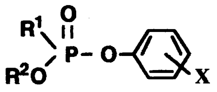
Class 6



Class 7



Class 8



Class 9

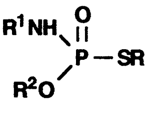
Class	#	Name	R¹	R²	X	Obsd.[a]	Calc.[b]	CLOGP
(6)	**34**	Butamifos	s-C_4H_9	C_2H_5	3-CH_3-6-NO_2	4.62	4.103	3.865
	35	Isofenphos	i-C_3H_7	C_2H_5	2-COO-i-C_3H_7	4.12	4.683	4.046
(7)	**36**	Oxon of (**34**)	s-C_4H_9	C_2H_5	3-CH_3-6-NO_2	2.73[c]	2.754	2.325
	37	Crufomate	CH_3	CH_3	2-Cl-4-t-C_4H_9	3.42	2.925	3.326
	38	Fenamiphos	i-C_3H_7	C_2H_5	3-CH_3-4-SCH_3	3.23	3.220	2.604
	39	ET-15[d]	H	CH_3	2,4,5-Cl_3	2.53	1.973	2.545
(8)	**40**	Cyanofenphos	Ph	C_2H_5	4-CN	4.29	4.152	4.283
	41	EPN	Ph	C_2H_5	4-NO_2	3.85	4.534	4.568
	42	Leptophos	Ph	CH_3	4-Br-2,5-Cl_2	6.31	5.846	6.390
	43	Trichloronate	C_2H_5	C_2H_5	2,4,5-Cl_3	5.23	4.835	4.939
(9)	**44**	Oxon of (**40**)	Ph	C_2H_5	4-CN	2.44[c]	3.095	2.473
	45	Oxon of (**42**)	Ph	CH_3	4-Br-2,5-Cl_2	4.58	4.502	4.580

[a,b]See footnotes of Table I-1. [c]Determined by the shake-flask method in our laboratory. [d]ET-15:*O*-methyl *O*-(2,4,5-trichlorophenyl) phosphoramidate.

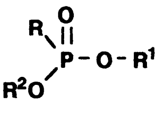
single P-B (B: O, NH, S or C) bonds by stepwise changing of the torsional angles. In the case of dimethyl phenyl phosphorothioate, which is the basic structure of class (1) compounds listed in Table I-1, the three lower-energy state conformers were found through the conformational analysis followed by full optimization of the geometry. The two conformers in which one of the two methyl groups is in the *trans* position to sulfur and the other is in the *gauche* position exhibited an energy about 1 - 2 kcal mol^{-1} lower than the other possessing both methyl groups in the *gauche* position. In each case, the phenyl group was in the *gauche* position. These conformations are found typically for the crystall structure of phosphorothionates (*28*). Similar stable conformations were expected for pesticides used here and the lowest-energy conformer with the extended structure was chosen for analysis.

The computational results are briefly summarized in Table III. The X-ray crystallographic structure of phosphorothioates and trisubstituted phosphates (I and II) were satisfactorily reproduced. The P-O bond length was slightly overestimated by 0.12 - 0.14Å but the deviations for the other bonds were less than 0.05Å. The mean unsigned error for bond angles was within 5°. The torsional angles determining the orientation of alkyl and aryl substituents were not predicted so precisely as the above geometric parameters and the mean unsigned error was less than 20°. The bond lengths in phosphoramidothioates (III, not shown), phosphoramidates (IV), and phosphonothioates (V, not shown) were estimated with a precision similar to those in phosphorothioates. The mean unsigned errors of bond and torsional angles were less than 4° and 17°, respectively. In the case of phosphorodithioates (VI), the P-S-C bond angles were overestimated by 15.6±0.9° (mean unsigned error, not shown) but the other geometric parameters were well reproduced. The S-P=S moiety has not

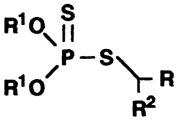
Table I-5. Structure and Log*P* Value of Organophosphorus Compound (Classes 10 - 13).



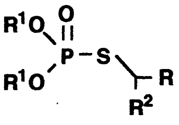
Class 10



Class 11



Class 12



Class 13

Class #	Name	R^1	R^2	R	Obsd.[a]	Calc.[b]	CLOGP
(10) **46**	Methamidophos	H	CH_3	CH_3	-0.66	-0.408	-0.868
47	DMPAT[c]	CH_3	CH_3	CH_3	-0.07	-0.062	-0.172
48	Acephate	$COCH_3$	CH_3	CH_3	-0.85	-0.648	-0.892
(11) **49**	MDMP[c]	CH_3	CH_3	CH_3	-0.66	-0.900	-0.662
50	Fosetyl	H	C_2H_5	H	-2.70	-2.623	N.E.[d]
51	Glyphosate	H	H	CH_2NHCH_2-COOH	-3.60	-4.085	N.E.[d]
52	Trichlorfon	CH_3	CH_3	$CH(OH)CCl_3$	0.51	0.737	0.299
(12) **53**	Dimethoate	CH_3	H	$CONHCH_3$	0.78	0.733	0.752
54	Azinphos-methyl	CH_3	H	H^{1e}	2.75	2.564	2.685
55	Phosmet	CH_3	H	H^{2e}	2.78	1.983	2.813
56	Phorate	C2H5	H	SC_2H_5	3.56	3.813	3.465
57	Formothion	CH_3	H	CON(CH3)CHO	1.48	1.103	1.507
58	Methidathion	CH_3	H	H^{3e}	2.42	2.373	2.409
59	Phenthoate	CH_3	Ph	$COOC_2H_5$	3.69	4.755	3.580
60	Terbufos	C_2H_5	H	S-*t*-C_4H_9	4.48	4.487	4.173
(13) **61**	Oxon of (**54**)	CH_3	H	H^{1e}	0.78	1.358	1.406
62	Oxon of (**56**)	C_2H_5	H	SC_2H_5	2.07	2.512	1.986

[a,b] See footnotes of Table I-1. [c] DMPAT: O,S-dimethyl methyl phosphoramidothioate, MDMP:O,O-dimethyl methyl phosphonate. [d] Not estimable by CLOGP. [e] Structures are shown below.

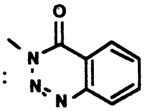
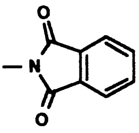
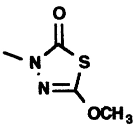
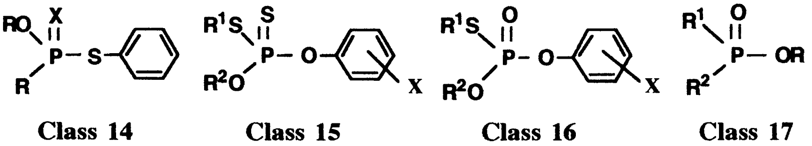
H^1:  , H^2:  , H^3: 

Table I-6. Structure and LogP Value of Organophosphorus Compounds (Classes 14 - 17).



Class #	Name	R	R¹	R²	X	Obsd.[a]	Calc.[b]	CLOGP
(14) **63**	Fonofos	C_2H_5			S	3.94	3.712	3.905
64	Oxon of (**63**)	C_2H_5			O	2.11	2.456	2.080
(15) **65**	Sulprofos		C_3H_7	C_2H_5	4-SCH_3	4.90	5.352	4.823
(16) **66**	Profenofos		C_3H_7	C_2H_5	4-Br-2-Cl	4.68	4.401	4.662
(17) **67**	MEPHOS[c]	C_2H_5	CH_3	H		-0.60	-1.417	-0.599

[a,b]See footnotes of Table I-1. [c]*O*-ethyl methyl phosphinate.

been considered in the atomic parametrization of phosphorus and sulfur in MNDO-PM3, which might result in the large deviations observed. These results showed that MNDO-PM3 gives the molecular geometries for organophosphorus compounds close to the crystallographic structures with a satisfactory precision.

Multiple Regression Analysis

A variety in the aryloxy moiety of phosphorothioates and phosphates (Classes 1 and 2) is considered to clarify the effect of electronic as well as steric properties on log*P*. The various connection patterns at phosphorus would be useful for estimating the effect of polarization. The following descriptors were taken into account for the analysis. The van der Waals volume (V) and surface area (S) were primarily chosen as steric descriptors since they are known to correlate well with log*P* (*29*). It has been considered that the partition coefficient of molecules is mostly determined by an extent of solvation and hence the solvent-accessible surface area (S_A) and volume (V_A) are the suitable descriptors (*7-9*). Therefore, these values were calculated on the basis of the PM3-optimized geometries.

As electronic factors, dipole momemt (D) and energy levels of the highest occupied and lowest unoccupied MOs (E_{HOMO} and E_{LUMO}) were calculated using MNDO-PM3. According to the Born expression (*13*), the electrostatic energy change in solvation was considered to approximately correlate with the summation of the squares of the net charge on all atoms ($\sum Q^2$). Because these descriptors alone represent a part of solvation energy, the thermodynamical descriptor accounting for the free energy change of solvation in total was further considered. If the free energy of a molecule in octanol is assumed to be approximated by that in the gas-phase, the free energy change in partitioning of the molecule from octanol to water (E_{OW}) can be estimated and easily transformed into log*P*. The PM3-SM3 method was used to

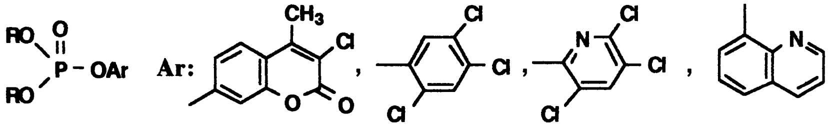
Table II. Organophosphorus Pesticides used for Comparison of the Molecular Geometry Between MNDO-PM3 and X-ray Crystallography.[a]

(I) Phosphorothioates (Classes 1, 3)

Methyl parathion (**#4**), Bromophos (**#5**), Iodofenphos (**#8**) (Class 1, Table I-1)
Chloropyrifos-methyl (**#24**), Chloropyrifos (**#26**) (Class 3, Table I-2)

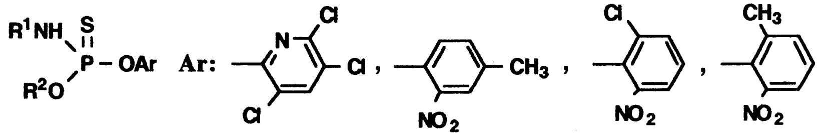
(II) Phosphates (Classes 2, 4)

Stirofos (**#30**) (Class 2, Table I-3)


RO
O
P – OAr
RO
Ar:
CH3
Cl
O
O
Cl
Cl
Cl
Cl
N
Cl
Cl
Cl
N

Coumaphos oxon, Ronnel oxon, Fospirate, Dimethyl 8-quinolyl
($R = C_2H_5$) ($R = CH_3$) ($R = CH_3$) phosphate ($R = CH_3$)

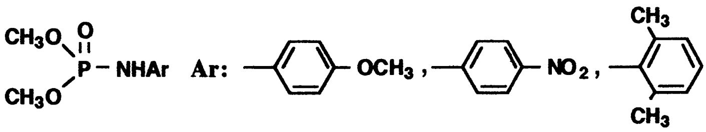
(III) Phosphoramidothioates (Class 6)


R1NH
S
P – OAr
R2O
Ar:
Cl
N
Cl
Cl
CH3
NO2
Cl
NO2
CH3
NO2

	IPAT	Amiprophos-M		
R^1 =	i-C_3H_7	i-C_3H_7	s-C_4H_9	s-C_4H_9
R^2 =	CH_3	C_2H_5	C_2H_5	C_2H_5

(IV) Phosphoramidates (Class 7)

N-Phenyl Derivatives:


CH3O
O
P – NHAr
CH3O
Ar:
OCH3
NO2
CH3
CH3

Crufomate (**#37**) (Table I-4)

(V) Phosphonothioates (Class 8)

EPN (**#41**), Leptophos (**#42**) (Table I-4)

(VI) Phosphorodithioates (Class 12)

Dimethoate (**#53**), Azinphos-methyl (**#54**) (Table I-5)
Amidithion: $R^1 = CH_3$, $R^2 = H$, $R = C(=O)NHCH_2CH_2OCH_3$

---

[a] Compounds without number were used only for the geometry comparison.

Table III. Comparison of Molecular Geometries Obtained by MNDO-PM3 and X-ray Crystallography.

	Mean unsigned error (deviation)/sign[a]			
Geometry	I. Phosphorothioates Classes 1 and 3 (n=5)[b]	II. Phosphates Classes 2 and 5 (n=5)[b]	IV. Phosphoramidates Class 7 (n=4)[b]	VI. Phosphorodithioates Class 12 (n=3)[b]
Bond length[c]				
P=A (A: O or S)	0.027 ± 0.004 /+	0.028 ± 0.037 /±	0.008 ± 0.005 /+	0.025 ± 0.006 /+
P-Oalkyl	0.134 ± 0.012 /+	0.132 ± 0.009 /+	0.135 ± 0.012 /+	0.136 ± 0.009 /+
P-Oaryl	0.124 ± 0.010 /+	0.140 ± 0.012 /+	0.120 ± 0.010 /+	
P-Balkyl (B: S or NH)			0.128 ± 0.012 /+	0.086 ± 0.010 /+
Bond angle[c]				
<APOalkyl	3.7 ± 2.5 /+	2.3 ± 1.2 /+	4.8 ± 1.1 /+	3.2 ± 0.6 /+
<APOaryl	4.9 ± 1.3 /+	2.4 ± 1.2 /±	3.7 ± 1.0 /+	
<APBalkyl			1.5 ± 1.0 /+	0.9 ± 0.9 /±
Torsional angle[c]				
<APOC(alkyl)	5.3 ± 3.3 /±	9.5 ± 4.8 /−	10.9 ± 9.0 /±	9.6 ± 9.3 /±
<APOC(aryl)	16.4 ± 7.0 /±	19.7 ± 14.6 /+	9.1 ± 5.2 /±	
<APBC(alkyl)			3.1 ± 1.6 /−	3.3 ± 1.6 /±
<POCC(aryl)	18.6 ± 12.7 /±	12.1 ± 10.8 /±	16.3 ± 5.8 /±	

[a] Sign; +, the larger value was obtained by MNDO-PM3 than X-ray crystallography; -, the opposite result to "+"; ±, no clear trend was detected. [b] Number of compounds tested. [c] Bond length in Å; bond and torsional angles in deg.

calculate the molecular properties of hydrated pesticides to obtaine the E_{OW} value. Since the geometric optimization of a molecule in the aqueous phase is known not to cause a significant change in the free energy of solvation (*30*), the optimized structure in the gas-phase was conveniently used for the PM3-SM3 calculations.

The effect of a hydrogen bond on solvation is not clearly taken into account in above descriptors. The hydrogen bonding site of a molecule is considered to possess a larger surface tension than the other sites and this difference should be considered (*8,19*). Although several quantum chemical methods have been introduced for this purpose (*6*), a simple idea of the total number of hydrogen bonds, NW, possibly formed for a functional group was taken in our approach basically according to Wolfenden's concept (*31*). This concept has stemed from the relation between the hydrophilic character of a compound and its vapor pressure over its aqueous solution. Based on the investigation on the partition coefficient of a compound (K*eq*) for transfer from water to a nonpolar environment such as an organic solvent, he further indicated that moderately polar groups in a molecule additionally influence K*eq*. Koehler et al. have developed this concept by using the MM2 molecular mechanics calculation on the hydrated molecules and have related the number of hydration for a functional group with Wolfenden's K*eq* (*8*). In this study, the donor (ND) and acceptor (NA) sites of the hydrogen bond were conveniently defined as listed in Table IV by referring to Koehler's approach (*8*). NW was defined as the summation of ND and NA. The ND value of nitrogen possessing a positive net charge was defined as zero, since it is unlikely to participate in the formation of the hydrogen bond. The NA value for P=O was defined to be 3 by comparing K*eq* values of phosphates (10^{-5} ~ 10^{-7}) (*32*) with Wolfenden's hydration potential scale (*8,31*). The transformation from P=O to P=S gave an increase of log$P$ by about two (Table I). A corresponding

Table IV. The Hydrogen-Bond Parameters for Each Functional Group.

Moiety	NW	ND	NA	Moiety	NW	ND	NA
Ar[R]SR	0	0	0	Dicarboximide	5	0	5
Ar[R]OR	0	0	0	Ar[R]NH_2	2	1	1
Ar[R]NO_2	1	0	1	Ar[R]NHR	2	1	1
Ar[R]COOR	1	0	1	Ar[R]OH	2	1	1
RCOOAr	3	0	3	Ar[R]COOH	3	1	2
Ar[R]C(=O)R	1	0	1	Ar[R]CONHR	4	1	3
Ar[R]CHO	1	0	1	Ar[R]$CONR_2$	3	0	3
Ar[R]CN	1	0	1	Ar[R]$CONH_2$	5	2	3
RS(=O)R	3	0	3	ArNHC(=O)R	3	1	2
R_3N	1	0	1	P=S	2	0	2
Ring-O[S]	0	0	0	P=O	3	0	3
Ring-N	1	0	1				

Ar: aryl, R: alkyl, ND: donor sites of the hydrogen bond, NA: acceptor sites of the hydrogen bond, NW: ND + NA.

100-fold change was expected for the partition from water to vapor which resulted in the decrease of *NA* by about 1 (*8*). Therefore, the *NA* value of P=S was defined to be 2. Since the S=O moiety was considered to be highly polarized, similar to P=O, the *NA* value was conveniently taken as 3 as the best fit realized in the multiple regression analysis. The *ND* value was defined to be 0 for P=O, P=S and S=O groups and *NW* = 0 when the moieties are not listed in Table IV except for Ar[R]SR and Ar[R]OR.

Results and Discussion

A high correlation was detected between log*P* and the steric descriptors. The single correlation coefficients for S, V, S_A, and V_A were 0.814, 0.862, 0.794, and 0.821, respectively. The partitioning has been described by the activity coefficient of a solute in each phase correlating with these descriptors (*7-9, 17-19*), which agrees with our results. The molecular volume was a better descriptor than the corresponding surface area and consideration of solvent-accessible surface did not improve the correlation. The previous studies (*9,28*) have shown that the relative significance among them is dependent on the electrostatic properties at the molecular surface used for analysis. The molecular surface area of the organophosphorus compounds consists of hydrophilic and hydrophobic sites each of which exhibits a different degree of solvation (*8,19,33*). The neglect of this difference in the calculation of S and S_A is likely to result in poorer correlations.

The other descriptors gave lower correlations, simple correlation coefficient being 0.410 with E_{OW}, -0.250 with $\sum Q^2$, 0.241 with D, 0.559 with E_{HOMO}, -0.561 with E_{LUMO}, and -0.629 with *NW*. The *NA* and *ND* descriptors afforded poorer correlations than *NW*. The unsigned simple correlation coefficient for each of the pairs among these and steric descriptors was less than 0.5. The two descriptors ($\sum Q^2$ and D) did not well describe log*P* by themselves and more detailed examinations to incorporate the electrostatic and polarization terms would be needed as previously reported (*11-13*). Although $\sum Q^2$ and D parameters are probably included in the E_{AV} descriptor, E_{AV} failed to show a good correlation with log*P*. The surface tension to hydration on each atom has not been parametrized for the organophosphorus compounds. The MNDO-PM3 method has shown to systematically give too positive a charge on nitrogen, as compared with the other MO methods (*25,29*). These two factors might result in the low correlation. The energy level of HOMO has been previously reported as one of the useful descriptors and is considered to describe the hydrogen-bond basicity of a solute molecule (*11,12*). In this study, *NW* showed the highest correlation with log*P* among the descriptors other than the steric parameters, followed by E_{HOMO} and E_{LUMO} with a similar correlation coefficient but the opposite sign. Therefore, the addition of *NW*, E_{HOMO} and E_{LUMO} descriptors to one of the steric parameters was expected to be useful to improve the regression equation describing log*P*.

Equations 1 - 4 were obtained by the regression analysis for 67 compounds when one more descriptor was added to each of the steric parameters.

$$\log P = 0.0353\ (\pm\ 0.0017)V\ -\ 0.5754\ (\pm\ 0.0466)NW\ -\ 1.4452\ (\pm\ 0.3839) \qquad (1)$$
$$r = 0.961,\ s = 0.546,\ F_{2,64} = 388.3$$

$$\log P = 0.0256\ (\pm\ 0.0015)S\ -\ 0.6396\ (\pm\ 0.0547)NW\ -\ 1.8268\ (\pm\ 0.4883) \qquad (2)$$
$$r = 0.945,\ s = 0.651,\ F_{2,64} = 264.6$$

$$\log P = 0.0101\ (\pm\ 0.0066)V_A\ -\ 0.6333\ (\pm\ 0.0528)NW\ -\ 2.3356\ (\pm\ 0.4936) \qquad (3)$$
$$r = 0.949,\ s = 0.627,\ F_{2,64} = 287.0$$

$$\log P = 0.0179\ (\pm\ 0.0012)S_A\ -\ 0.6505\ (\pm\ 0.0602)NW\ -\ 3.3481\ (\pm\ 0.6403) \qquad (4)$$
$$r = 0.932,\ s = 0.717,\ F_{2,64} = 212.3$$

NW is the only descriptor improving the correlation coefficient to >0.9, which was justified above 99.5% by the F test. The regression coefficient of NW was negative, indicating that the greater the extent of hydration of a molecule, the lower the log$P$.

Addition of E_{LUMO} was found to afford a better fit and the F test showed that further addition of the other descriptor terms did not improve the correlation coefficient. Equation 5 was the best to describe the log$P$ value of organophosphorus compounds included (n = 67).

$$\log P = 0.0314\ (\pm\ 0.0014)V\ -\ 0.5598\ (\pm\ 0.0363)NW\ -\ 0.4178\ (\pm\ 0.0639)E_{LUMO}$$
$$-\ 1.3764\ (\pm\ 0.2990) \qquad (5)$$
$$r = 0.977,\ s = 0.425,\ F_{3,63} = 441.7$$

The correlation between calculated and experimental log$P$ values is shown in Figure 1. LUMO is located at the P=S or P=O moiety of the pesticides and the orbital interactions with the oxygen lone pair of water and 1-octanol are likely to occur. The E_{LUMO} term in equation 5 was negative and hence the decrease of the energy level of LUMO increases the log$P$ value. This profile suggests that the lower the LUMO of P=S and P=O moieties, the easier the interaction of the more basic hydroxyl oxygen lone-pair of 1-octanol leading the partition of the solute into the octanol phase. The unsigned errors for logP_{calc} and CLOGP3 values were 0.331 ± 0.249 (n = 67) and 0.213 ± 0.236 (n = 65), respectively. Although the precision in the estimation of log$P$ by our method is slightly poorer than that by the CLOGP procedure, equation 5 gave satisfactory log$P$ values for the two compounds which were not calculable by CLOGP due to the lack of the fragment values (Table I-5). The present method was found to satisfactorily estimate the wide range of the log$P$ values from -3.6 to 6.3 for the various organophosphorus pesticides. Furthermore, our method can afford valuable information on the geometry and electronic properties of the molecule in lower energy states which can not be done by CLOGP.

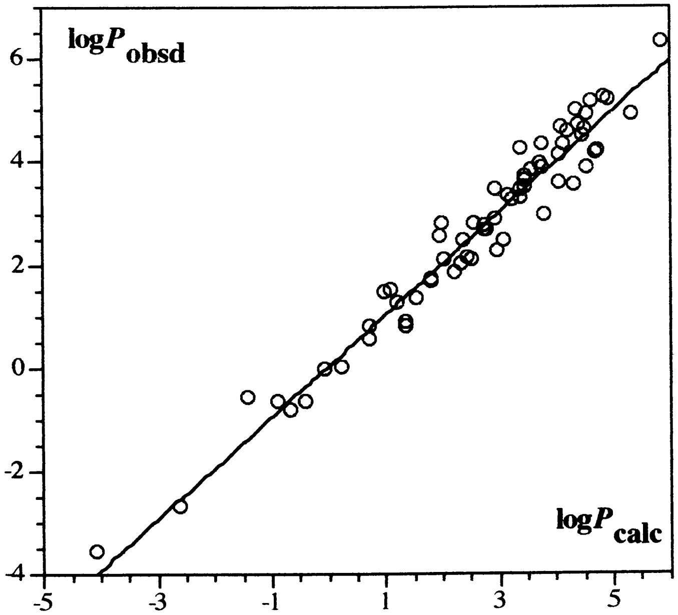


Figure 1. Correlation between Estimated ($\log P_{calc}$) and Experimental ($\log P_{obsd}$) $\log P$ Values.

Acknowledgments

We are most grateful to Professor Corwin Hansch of Pomona College for giving us the chance to write this chapter and also to Professor Toshio Fujita for the helpful suggestions throughout the writing.

Literature Cited

1. Klein, A. W. In *Handbook of Environmental Chemistry*; Hutzinger, O. Ed.; Springer-Verlag: Berlin, 1985, Vol. 2, Part C; pp 1 - 28.
2. Wasik, S. P.; Miller, M. M.; Tewari, Y. B.; May, W. E.; Sonnefeld, W. J.; DeVoe, H.; Zoller, W. H. *Residue Rev.* **1983**, *85*, 29.
3. Sicbaldi, F.; DelRe, A. A. M. *Rev. Environ. Contam. Toxicol.* **1993**, *133*, 59.
4. Leo, A. J. In *Comprehensive Medicinal Chemistry*, Hansch, C. Ed.; Pergamon Press: Oxford, 1990, Vol. 4; pp 295 - 319.
5. Chou, J. T.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 172.
6. Leo, A. J. *Chem. Rev.* **1993**, *93*, 1281.
7. Hermann, R. B. *J. Phy. Chem.* **1972**, *76*, 2754.
8. Koehler, M. G.; Grigoras, S.; Dunn III, W. J. *Quant. Struct.-Act. Relat.* **1988**, *7*, 150.
9. Bruijn, J.; Hermens, J. *Quant. Struct.-Act. Relat.* **1990**, *9*, 11.
10. Klopman, G.; Iroff, L. D. *J. Comput. Chem.* **1981**, *2*, 157.

11. Schüürman, G. *Quant. Struct.-Act. Relat.* **1990**, *9*, 326.
12. Lewis, D. F. V. *J. Comput. Chem.* **1989**, *10*, 145.
13. Cammarata, A.; Rogers, K. S. *J. Med. Chem.* **1971**, *14*, 269.
14. Leahy, D.; *J. Pharm. Sci.* **1986**, *75*, 629.
15. Leahy, D.; Morris, J. J.; Taylor, P. J.; Wait, A. R. *J. Chem. Soc. Perkin Trans. 2* **1992**, 705.
16. Hopfinger, A. J.; Battershell, R. D. *J. Med. Chem.* **1976**, *19*, 569.
17. Kasai, K.; Umeyama, H.; Tomonaga, A. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 2701.
18. Camilleri, P.; Watts, S. A.; Boraston, J. A.: *J. Chem. Soc., Perkin Trans II* **1988**, 1699.
19. Sasaki, Y.; Kubodera, H.; Matsuzaki, T.; Umeyama, H. *J. Pharmacobiol.-Dyn.* **1991**, *14*, 207.
20. Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.
21. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209 & 221.
22. Katagi, T. *J. Comput. Chem.* **1993**, *14*, 1250 and references therein.
23. Broyden, C. G. *J. Inst. Math. Appl.* **1970**, *6*, 222; Fletcher, R. *Comput. J.* **1970**, *13*, 317; Goldfarb, D. *Math. Comput.* **1970**, *24*, 23; Shanno, D. F. *Math. Comput.* **1970**, *24*, 647.
24. Pascual-Ahuir, J. L.; Silla, E.; Tomasi, J.; Bonaccorsi, R. *J. Comput. Chem.* **1987**, *8*, 778.
25. Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Design* **1992**, *6*, 629.
26. Biobyte Corporation, Claremont, CA.
27. Yoshida, M.; Takayama, C.; Morooka, S.; Yokota, A. *7th Int. Conf. Comput. Res. Educ.* No. 32, 1985.
28. Allen, F. H.; Johnson, O.; Macrae, C. F.; Smith, J. M.; Motherwell, W. D. S.; Gally, J. J.; Watson, D. G.; Rowland, R. S.; Edgington, P. R.; Garner, S. E.; Davies, J. E.; Mitchell, G. F. *The Cambridge Structural Database, CSD System Documentation*; Cambridge Crystallographic Centre: Cambridge, U.K., 1992.
29. Moriguchi, I.; Kanada, Y.; Komatsu, K. *Chem. Pharm. Bull.* **1976**, *24*, 1799.
30. Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089.
31. Wolfenden, R. *Science* **1983**, *222*, 1087.
32. Wolfenden, R.; Williams, R. *J. Am. Chem. Soc.* **1983**, *105*, 1028.
33. Richards, N. G. J.; Williams, P. B.; Tune, M. S. *J. Quant. Chem., Quant. Biol. Symp.* **1991**, *18*, 299.

RECEIVED April 26, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Chapter 5

The Calculation of Pesticide Hydrophobicity by Computer

Albert J. Leo

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711–6338

Estimation of hydrophobicity of a prospective pesticide structure is often desireable prior to its synthesis. This parameter, as log P_{oct}, has been useful in predicting translocation in plants as well as in predicting bioaccumulation in animals and transport in soils, etc. (*1-3*) This report gives some CLOGP (*4*) calculations for pesticides containing a variety of toxiphores: carbamates, DDT analogs, sulfonylureas, methoxyureas, and various phosphorous-containing cholinesterase inhibitors.

Carbamate-Based Pesticides

Almost all of the pesticides with the 'true' carbamate moiety are calculated well, as seen in Table I. So too are most of those where the carbamate moiety is highly modified, such as methomyl (#18). Some of the structures are depicted in Fig. 1. A significant positive deviation occurs when the carbamate is substituted on a very electron-deficient ring, such as on pyrimidine in pirimicarb (#11). A deviation of opposite sign results in the case of carbendazim's benzimidazole ring (#15), but not with benomyl (#8) in which the imidazole nitrogen is substituted with a carboxamido group. The antineoplastic, diaziquone (#22) was included merely to show that a very complex structure containing two carbamate groups is also calculated well. Note also that in methomyl CLOGP uses the value for a fragment larger than the simple carbamate, which is also true for alanycarb and asulam.

When anomalies are encountered in the calculation of log P_{oct}, it may be an indication of something of greater importance than the solute's unusual hydrophobicity. Some recent measurements by Takahashi et al. (*5*) indicate that the carbamate group may take an unexpected conformation in water, allowing a 3-ethoxy substituent to interact with the ester portion of the carbamate moiety in some sort of 'hydrophobic overlap'. Their data for some 3,4-diethoxy analogs are shown together with some mono-substituted analogs in Table II.

It will be noted that when the phenyl ring contains no more than one small substituent (#1-6), the alkyl ester group, R, can be as large as pentyl without encountering calculation anomalies. When 3,4-diethoxy groups are present, however,

0097–6156/95/0606–0062$12.00/0

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Table I. Log P_{oct} of Carbamate-Based Pesticides

No.	*Name*	*Meas.*	*CLOGP*	*Diff.*
1.	phenmedipham	3.59	3.52	+0.07
2.	chlorpropham	3.51	3.52	-0.01
3.*	alanycarb	3.43	3.49	-0.06*
4.	desmedipham	3.39	3.55	-0.16
5.	methiocarb	2.92	2.87	+0.05
6.	carbaryl	2.36	2.38	-0.02
7.	carbofuran	2.32	2.32	0.00
8.*	benomyl	2.12	1.94	+0.18*
9.	meobal	2.09	2.16	-0.07
10.	thiodicarb	1.70	1.70	0.00
11.*	pirimicarb	1.70	1.18	+0.52*
12.*	bendiocarb	1.70	1.67	+0.03*
13.*	karbutilate	1.66	1.47	+0.19*
14.	metoxuron	1.64	1.85	-0.21
15.	carbendazim	1.52	1.86	-0.31
16.	baygon	1.52	1.65	-0.13
17.	aldicarb	1.13	1.12	+0.01
18.*	methomyl	0.60	0.54	+0.06*
19.*	asulam	-0.27	-0.31	+0.04*
20.	oxamyl	-0.47	-0.47	0.00
21.	aldicarb-sulfone	-0.57	-0.42	-0.15
22.*	diaziquone†	-0.02	-0.10	+0.08*

*Structure shown in Fig. 1.; †Antineoplastic Av. Diff. = ±0.107

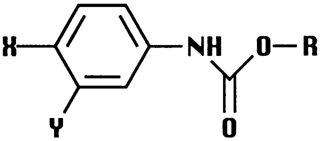


Table II. Log P Anomalies with N-Phenyl Carbamates

No.	*R*	*X*	*Y*	*Meas.*	*CLOGP*	*Dev.*
1	Me	H	H	1.76	1.76	0.00
2	Me	H	OMe	1.92	1.81	+0.11
3	Me	OMe	H	1.69	1.81	-0.12
4	Et	H	H	2.30	2.29	+0.01
5	Pr	H	H	2.80	2.82	-0.02
6	Pent	H	H	3.80	3.88	-0.08
7	Me	OEt	OEt	2.09	2.09	0.00
8	Et	OEt	OEt	2.50	2.62	-0.12
9	Pr	OEt	OEt	2.91	3.15	-0.24
10	i-Pr	OEt	OEt	2.82	2.93	-0.11
11	CH_2Ph	OEt	OEt	3.29	2.86	-0.57
12	$CH(CH_3)CH_2F$	OEt	OEt	2.59*	2.66	-0.07
13	$CH_2CH_2OCH_3$	OEt	OEt	2.31	1.92	+0.39
14	$CH(CH_3)CH_2OCH_3$	OEt	OEt	2.56*	2.23	+0.33
15	$CH_2CH_2C{\equiv}N$	OEt	OEt	2.13	1.57	+0.56

*Values from HPLC regression

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

a successively larger negative correction is required if R is larger than methyl. It amounts to -0.24 for propyl and -0.57 for benzyl. This negative correction is postulated to be required due to an overlap between the hydrocarbon portions of the ethoxy and the ester R groups which occurs in the aqueous phase.

Figure 2. shows the bond rotations which would allow the two hydrophobic portions of the diethoxyphenyl carbamates to be in a position where a part of their surface would *not* be solvated by water. Previous work has shown that lone pair repulsion would keep ortho-methoxy groups from being planar with the ring with the result that these substituents appear more hydrophilic than usual. It was assumed that the same negative correction factor would apply to all ortho-alkoxy substituents. However, the report by Takahashi et al (*5*) prompted our laboratory to measure 1,2-diethoxy benzene and 3,4-diethoxy benzoic acid. We learned that the negative correction for the ortho-diethoxy moiety is nearly twice as large as for the dimethoxy. This information is included in the CLOGP values in Table II, but is not yet coded in the released versions of the program.

Rotation of the carbamate bond out of the plane of the phenyl ring does not 'cost' a great deal of energy, and CONCORD (*6*) shows it at 90° as seen in Fig.2a. Fig. 2b shows the bond rotations needed for maximum hydrophobic overlap which would result in a decreased solute size and favor water solvation (*7*). The upper space filling depiction in Fig. 3 shows how the O-propyl analog would look in the extended configuration and the lower depicts it is in the overlapped conformation. The O-benzyl analog, depicted as space-filling in Fig 4a, shows an even tighter overlap, which is consistent with its greater negative deviation. (See Table II, #11) This 'overlap hypothesis' suggests that the O-butyl analog (the upper structure in Fig. 4b) would fold toward the 3-ethoxy, but the methoxyethyl analog of the same length (lower structure in Fig. 4b) would not, because that would push the polar oxygen against a hydrocarbon surface. On the contrary, the impetus for folding may be in the opposite direction where an internal H-bond could form with the ether oxygen as acceptor. Indeed, as seen for solutes #13-15 in Table II, R-groups with the 2-ether oxygen need a positive rather than a negative correction, because internal H-bonds between acyclic groups are not yet recognized by CLOGP.

Additional support that 'hydrophobic overlap' might compensate for an otherwise unfavorable bond rotation is seen in a set of O-bornyl-2-alkoxyphenylcarbamates recently reported by Gregan et al. (*8*) and shown in Fig. 5a and depicted as space-filling in 5b. When the same carbamate bond rotations shown previously in Fig. 2 are allowed, then an R-group longer than butyl can come in close contact with the bornyl moiety. The normal log P increment for a methylene group (about 0.5) is seen from methyl through butyl in their report, but it drops to 0.31 for longer chains. With the usual planar carbamate, CONCORD shows the pentyloxy group close to the amine nitrogen, but Gregan's measurements were taken at a pH where the nitrogen would be protonated, and that makes it even more likely that the anilino group would be rotated so that its pentyloxy substituent could overlap the bornyl's hydrocarbon surface.

Interaction between topologically distant groups in these carbamate solutes seems impossible without some bond rotations which would ordinarily be unfavorable if it were not for strong solvation forces. The 'hydrophobic overlap' hypothesis certainly needs confirmation by other physico-chemical measurements before it is accepted with confidence.

Phosphorous-Containing Pesticides

With few exceptions, CLOGP calculates phosphate, phosphonate and phosphoramidate pesticides and their thio analogs quite satisfactorily, as seen in Table III. The measured log P values of these examples range from zero to 5.3, and, since most are taken from the literature and from a number of laboratories, a reasonable estimate of error in measurement would be ± 0.06. Except for chlorpyrifos (#5) the average calculation error is little more than twice this amount, that is, 0.146. Note that if the 'slow-stir' value for chlorpyrifos were ignored, it would cut the deviation for that solute in half.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

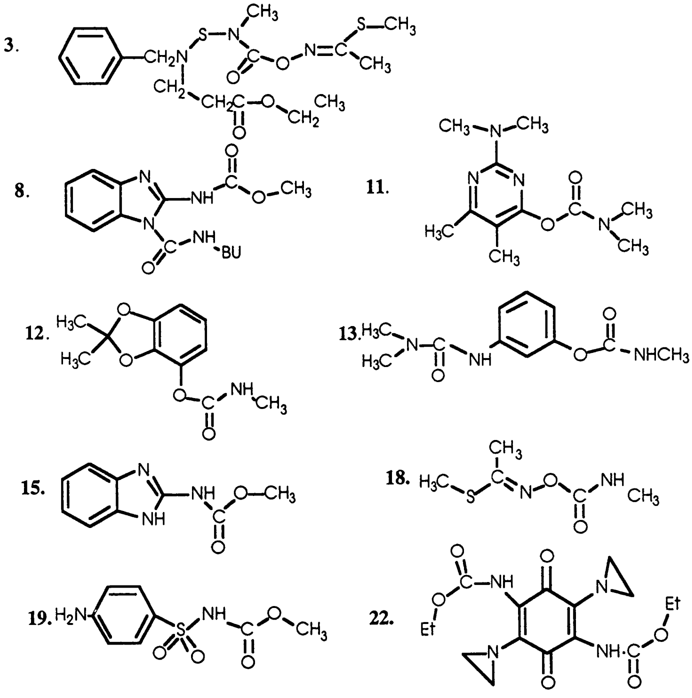
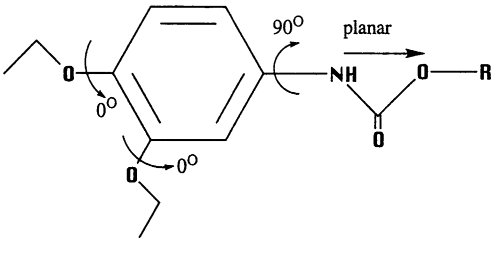

3.
8.
11.
12.
13.
15.
18.
19.
22.

Fig. 1 Structures of some of the pesticides in Table 1.

(a) CONCORD Co-ordinates


90°
planar
0°
0°
NH
O
R
O
O
O

(b) Folded Conformation

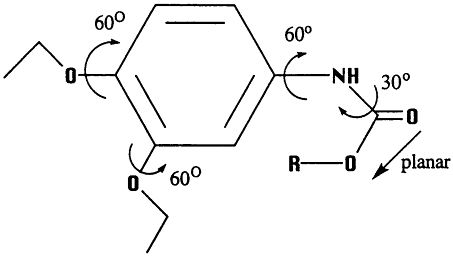

60°
60°
60°
30°
NH
O
O
O
R
O
planar

Fig. 2 Possible conformations of 3,4-diethoxyphenyl carbamates

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

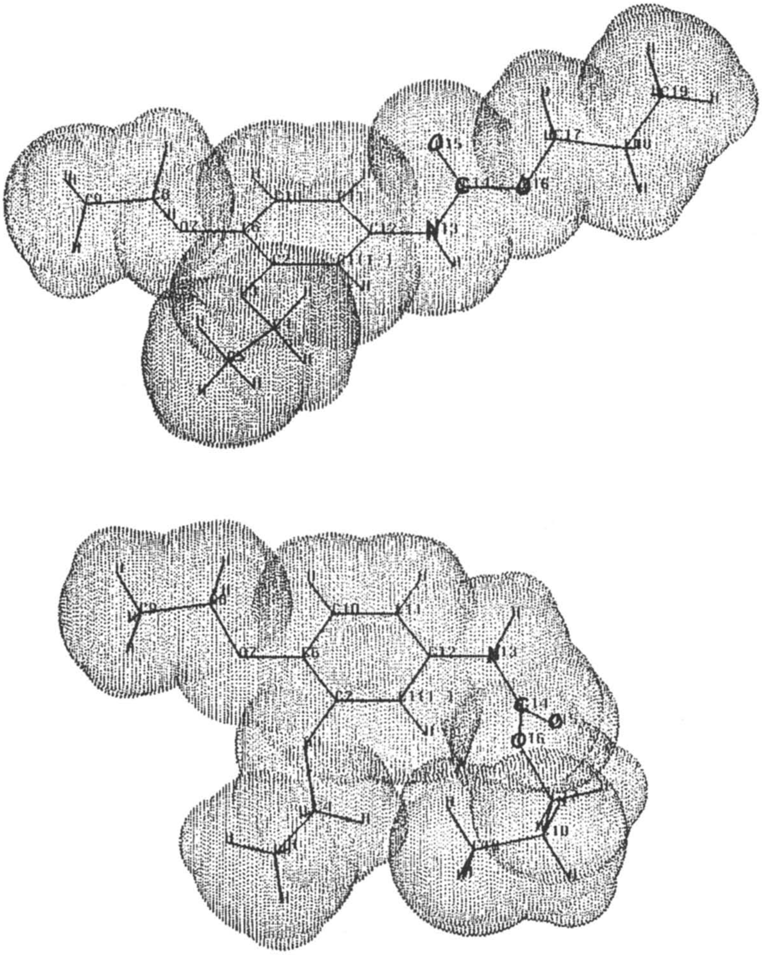


Fig. 3 Extended (upper) and folded (lower) conformations of Cpd. #9 of Table 2.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

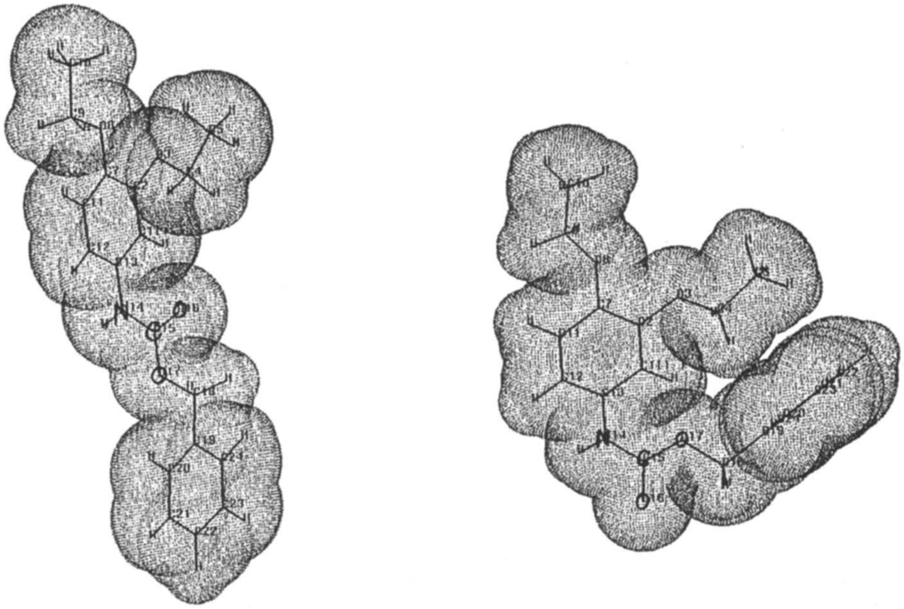


Fig. 4a. Extended and folded conformations of Cpd. #11 of Table 2.

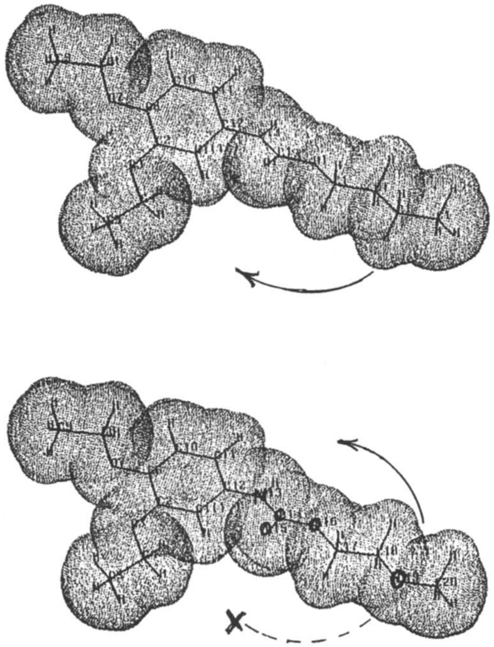


Fig. 4b. Carbamates of Table 2: upper, R = butyl; lower, R = methoxyethyl (#13)

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

O-Bornyl-(2-Alkoxyphenyl)carbamates

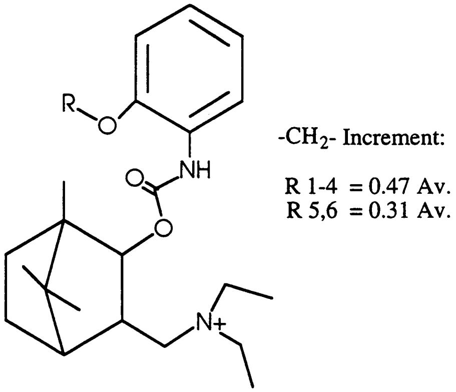

R
O
NH
O
O
N+
-CH$_2$- Increment:
R 1-4 = 0.47 Av.
R 5,6 = 0.31 Av.

Fig. 5a Overlap in O-(Methylbornyl)-2-alkoxyphenyl carbamates when R greater than butyl.

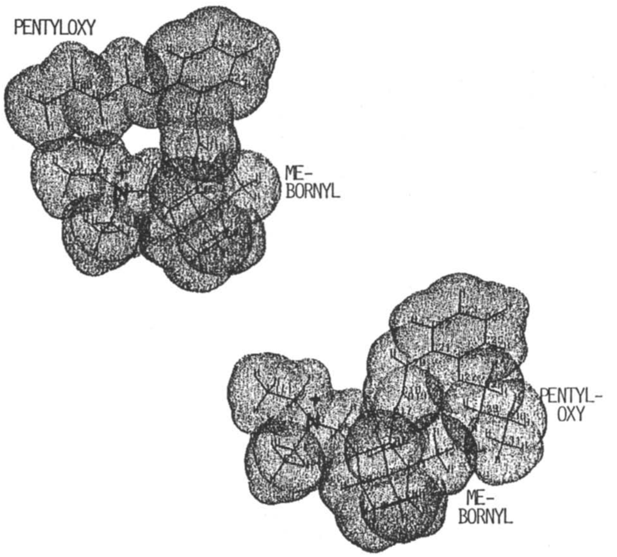

PENTYLOXY
ME-
BORNYL
PENTYL-
OXY
ME-
BORNYL

Fig.5b. Space-filling depiction of O-bornyl-phenyl carbamates.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Table III. Log P of Phosphorous-Containing Pesticides

No.	*Name*	*Meas.*	*CLOGP*	*Diff.*
1	carbophenothion	5.33	5.50	-0.17
2	parathion	3.83	3.47	+0.36
3	dimethylparathion	2.86	2.79	+0.07
4	paraoxon	1.98	2.06	-0.08
5	chlorpyrifos	5.27*	4.44	+0.83
6	malathion	2.36	2.40	-0.04
7	diazinon	3.81	3.50	+0.31
8	dicrotophos	0.00	0.40	-0.40
9	dimethoate	0.78	0.75	+0.03
10	fenthion	4.09	3.91	+0.18
11	isofenphos	4.12	4.24	-0.12
12	methidathion	2.42	2.41	+0.01
13	pebulate	3.84	3.74	+0.10
14	fonofoxon	2.11	2.08	+0.03
	Average difference	=		**±0.15**

*Measured by 'slow-stir'; other shake-flask value = 4.82.

DDT Analogs

As expected, DDT analogs are very hydrophobic, and they present considerable difficulties in measurement. Some of the early measurements using the "shake-flask" technique with radio-labelled material (*9*) were almost three log units lower than those calculated by CLOGP and were clearly in error. Solutes having a true log P greater than six can be easily 'emulsified' in the aqueous phase by octanol, especially if vigorous shaking incorporates more than the 10^{-3} M normally present at saturation. As the "slow-stir" technique became established (*10*), reproducible values were obtained which are much more in line with calculation, as seen in Table IV.

Table IV. DDT and Analogs

No.	*Name*	*Meas.*	*CLOGP*	*Diff.*
1.	DDE	6.96	6.94	+0.02
2.	DDT	6.91	6.76	+0.15
3.	Methoxychlor	5.08	5.17	-0.09
4.	Dicofol	4.28	6.06	1.78

The measured value for dicofol (*11*) is so low that one is tempted to write it off as experimental error. This conclusion seems supported by the fact that, as seen fin Fig. 6, the difference in log P_{oct} between 1,1,1-trichloroethane (2.49) and trichloroethanol (1.42) is only 1.07 compared to the 2.63 difference between DDT and dicofol. However, one should not overlook the fact that the unexpectedly low log P_{oct} of dicofol may be indicative of an important difference between the solvation properties of the organic and aqueous phases; namely, that it may be difficult for octanol as an H-donor to access the hindered hydroxyl in dicofol while it would have no such difficulty in trichloroethanol. Hindered access by the solvent, octanol, has previously been postulated to explain the unexpectedly low values for phenols and pyridines with bulky substituents in the 2 and 6 positions (*4*). In the case of 2,6-bis-(sec-butyl)phenol, the difference between measured and calculated log P is -1.03 log units.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Hindered Octanol Access to H-Bonding Groups (?)

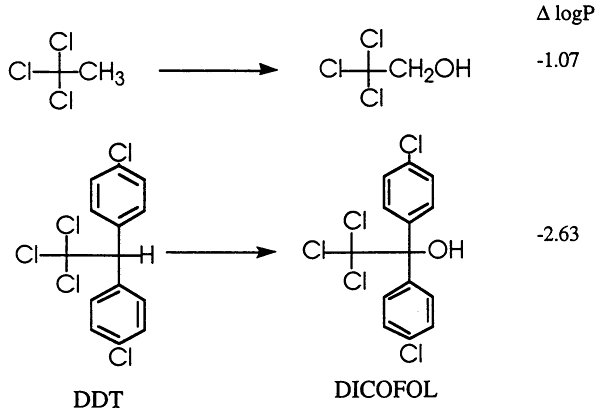
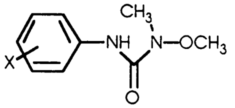

Δ logP
Cl
Cl
CH3
Cl
Cl
Cl
CH2OH
Cl
-1.07
Cl
Cl
Cl
H
Cl
Cl
Cl
Cl
Cl
OH
Cl
Cl
-2.63
DDT
DICOFOL

Fig. 6. Hindered access by octanol to H-bonding groups

1-Methoxy-ureas: Linuron Analogs


CH3
X
NH
N—OCH3
O

Name	X	Meas.	CLOGP	Δ
Linuron	3,4-Cl2	3.20	3.00	+.20
Monlinuron	4-Cl	2.30	2.31	-.01
Trifluron	3-CF3	2.64	2.71	-.08
4-Benzyloxy-		3.11	3.17	-.06

27 analogs: Av. Δ = 0.12

Oxy-urea fragment assigned same electronic susceptibility, rho, as urea: 1.08.

No ortho substituted analogs. No interaction with other ring substituents as with Carbamates.

Fig. 7. N-methoxy ureas

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Urea-type Herbicides

Methoxyureas Dependable measured log P_{Oct} values are available for 29 methoxyurea herbicides, a few of which are shown in Fig. 7. CLOGP calculates these with an average error of ±0.18. Removing just two of these reduces the average error to ±0.12 which is about twice the error in measurement. Since none of these analogs have a substituent ortho to the urea group and all seem to show only a normal electronic interaction by the meta and para substituents, the good agreement seen between measured and calculated values is expected.

Sulfonylurea-types A few of the more common sulfonurea structures with measured log P values (*12*) are shown in Fig. 8, along with a sulfonamido analog, Londax. Since electronic interactions make a considerable contribution to the calculated log P (3.43 for #4, Classic) the error level is lower than might be expected, especially considering the fact that the *sym*-triazine ring appears in the first three shown.

***sym*-Triazine-Based Types** The *sym*-triazine-based herbicides present quite a challenge to calculation of log P_{Oct}. The parent sym-triazine is not very stable and one cannot rely very heavily on it's measured value of -0.73 (*13*), which is +0.57 higher than calculated by CLOGP. Fortunately the more stable triazine herbicides are better behaved: 62 analogs are calculated within ±0.4 of the measured value. However, the calculation algorithm must take into account the fact that the electronic interactions are not additive in the usual Hammett sigma sense. In Fig. 9 it should be noted that the calculation for simazine shows how a scaling down of all *potential* electronic interactions must be made to produce a reasonable estimate. The estimations for the benzoguanamines and branched alkyl melamines are much poorer, however. These seem to require a type of steric correction which will necessitate a major code revision in CLOGP.

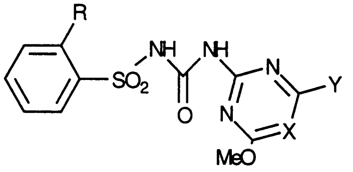
Pyrrole-Dicarboxylates This set of 5'-(chloropyridyl)-pyrrole-3',4'-dicarboxylates described in 1990 by Andrea et al (*14*) provides a good illustration of how difficult it can be to quantify the effect steric interactions can have on log P. As seen in Fig. 10, CLOGP calculates all but the 3-Cl analog with acceptable precision. All contain a correction factor of -0.60 for the steric decoupling of the 3',4'-dicarboxylate groups on the pyrrole ring. The paper by Andrea et al points out that if only the pyrrole ring were considered, electronic structure analyses and computer graphics would allow either of two conformations: both carboxylate group twisted at the same angle to the ring plane, or one remaining in the plane and the other twisted at 90°. The CLOGP factor of -0.60 might be appropriate for either case.

With no substituent at the 3-position of pyridine, the maximum resonance stabilization would be achieved with the 4'-carboxylate planar and the 3' perpendicular. However, for the 3-chloropyridyl analog (and also the 1-isoquinolyl), the 4'-carboxylate must be perpendicular while the 3' is planar. CLOGP does invoke a -0.25 correction for the Cl ortho to a second aromatic ring, a value established by numerous 2,2',6,6'-PCB analog (*1*). But the currently exported version of the algorithm cannot deal with the need for a *further* -1.0 log correction caused by the carboxylate decoupling in these herbicides. It is in the 'in-house' development version of CLOGP, but further confirmation of the generality of such a correction is needed before including it in the general release.

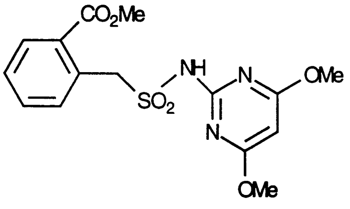
The 3-chloro analog exhibits higher pre-emergence activity than the 4 & 5, while the latter have higher post-emergence activity. The difference in log P, as it affects transport, may have some bearing on these observed activities, but other factors are undoubtedly involved.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Sulfonylurea-type Herbicides[1]



Name	R	X	Y	Meas.[2]	CLOGP
1. Allyl	CO_2Me	N	Me	1.70	1.80
2. Glean	Cl	N	Me	2.14	2.14
3. Harmony	CO_2Me	N	Me	1.52	1.49
	(on Thiophene-3-sulfonurea-))				
4. Classic	CO_2Et	CH	Cl	3.31	3.38



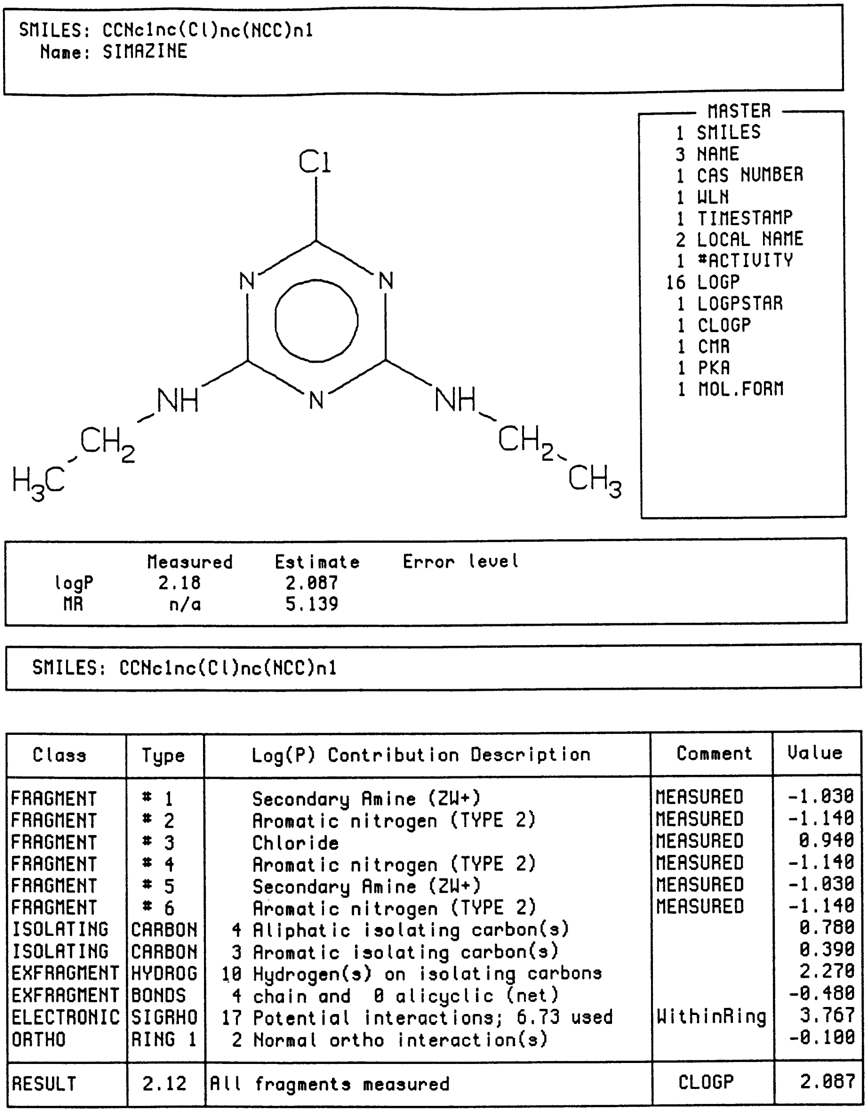
5. Londax 2.40 2.35

[1] J. Hay, Pestic. Sci., 29, 247 (1990)
[2] Measured at pH 5.0; corrected to neutral form.

Fig. 8 Sulfonurea-type herbicides

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005



```
SMILES: CCNc1nc(Cl)nc(NCC)n1
  Name: SIMAZINE
```

```
 ——— MASTER ———
  1 SMILES
  3 NAME
  1 CAS NUMBER
  1 WLN
  1 TIMESTAMP
  2 LOCAL NAME
  1 *ACTIVITY
 16 LOGP
  1 LOGPSTAR
  1 CLOGP
  1 CMR
  1 PKA
  1 MOL.FORM
```

```
          Measured    Estimate    Error level
  logP      2.18        2.087
  MR        n/a         5.139
```

```
SMILES: CCNc1nc(Cl)nc(NCC)n1
```

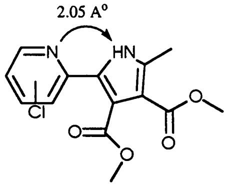
Class	Type	Log(P) Contribution Description	Comment	Value
FRAGMENT	# 1	Secondary Amine (ZW+)	MEASURED	-1.030
FRAGMENT	# 2	Aromatic nitrogen (TYPE 2)	MEASURED	-1.140
FRAGMENT	# 3	Chloride	MEASURED	0.940
FRAGMENT	# 4	Aromatic nitrogen (TYPE 2)	MEASURED	-1.140
FRAGMENT	# 5	Secondary Amine (ZW+)	MEASURED	-1.030
FRAGMENT	# 6	Aromatic nitrogen (TYPE 2)	MEASURED	-1.140
ISOLATING	CARBON	4 Aliphatic isolating carbon(s)		0.780
ISOLATING	CARBON	3 Aromatic isolating carbon(s)		0.390
EXFRAGMENT	HYDROG	10 Hydrogen(s) on isolating carbons		2.270
EXFRAGMENT	BONDS	4 chain and 0 alicyclic (net)		-0.480
ELECTRONIC	SIGRHO	17 Potential interactions; 6.73 used	WithinRing	3.767
ORTHO	RING 1	2 Normal ortho interaction(s)		-0.100
RESULT	2.12	All fragments measured	CLOGP	2.087

Fig. 9. CLOGP calculation for simazine

Conclusions

The correction factors needed to account for the non-additivity of hydrophobic fragmental values in log P_{Oct} calculation are not intended to be arbitrary indicator variables. Considerable effort is being made to relate them directly to solvatochromic parameters determined by other physico-chemical measurements, especially to α and β hydrogen bonding parameters of Kamlet, Taft et al (*15*). It is gratifying to see that some of the ortho factors incorporated into the CLOGP algorithm can be correlated with

5-(Chloropyridyl)-Pyrrole-3,4-dicarboxylates



	Log P_{oct}	
	Meas.*	CLOGP
3-Cl	1.51	2.49
4-Cl	2.63	2.74
5-Cl	2.70	2.74
6-Cl	2.65	2.69

CLOGP uses Factor of -0.60 for dicarboxylate twisting.

CLOGP does *not* need HB Factor between nitrogens*.

CLOGP uses Factor of -0.25 for twisting by 3-Cl; needs -1.25

*T. Andrea et al, Pestic Sci., 28, 49 (1990)

Fig. 10. 5-(Chloropyridyl)-pyrrole-3,4-dicarboxylates

decreased solvent-accessible-surface-area (SASA), as with the PCBs or, as noted above, with X-ray coordinates and/or quantum chemical calculations. In the future it is hoped that some of the simpler factors can be determined rapidly by M.O. or molecular mechanics programs so that they can be incorporated 'on-the-fly' in a higher-level version of CLOGP. But until then, the disagreement between measured and calculated log P values can often provide useful insights into conformational and electronic phenomena not readily determined by other means.

Literature Cited

1. Leo, A. in *QSAR in Environmental Toxicology-IV* Hermens & Opperhuizen Eds., Elsevier, Amsterdam, 1991, p.121.
2. Leo, A. *Methods in Enzymology* **1991**, 202-A, 544.
3. Hansch, C., *Accounts of Chemical Research* **1993**, 26, 147
4. Leo, A., *Chemical Reviews* **1993,** 93, 1281
5. Takahashi,J., Kirino,O., Takayama, C. and Kamoshita, K. *J. Chromatog.* **1988**, 436, 316.
6. CONCORD program written by R. Pealman, Univ. of Texas, Austin; furnished courtesy of Tripos Assoc., St. Louis, Mo.
7. Leahy, D. *J. Pharm. Sci.* **1987**, 75, 629.
8. Gregan,F. & Polasek,E. *Ceskosl. Farmacie* **1992**, 41, 303.
9. Metcalf, R. L., & Lu, P. Y., *E.P.A. Report* **1978**
10. De Bruijn,J, Busser, F., Seinen, W. & Hermens, J. *Environ.Tos. Chem.* **1989**, 8 499.
11. Fujita, T., EMIL project, **1992**, private communication.
12. Hay, J. Pestic. Sci., **1990**, 29, 247 .
13. Brugnoni, G., Moser,P. & Trebst, A. *Z. Naturforsch.* **1979,** 34, 1028 and private communication.
14. Andrea, T. A., Stranz, D. D., Yang, A., Kleier, D. A., Patel, K. M., Powell, J. E., Price, T. P. & Marynick, D. S. *Pestic. Sci.* **1990**, 28, 49.
15. Kamlet,M., Doherty, r., Abraham, M. and Taft, R. *Quant.Struct.-Act. Relat.* **1988**, 7, 71.

RECEIVED July 28, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 16, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch005

Chapter 6

Transport Parameter Dependence on Intermolecular Forces

Marvin Charton

Chemistry Department, Pratt Institute, DeKalb Avenue and Hall Street, Brooklyn, NY 11205

The intermolecular force (IMF) model provides a much better quantitative description of transport parameters for substituted benzenes and alkanes when the dipole moment μ is included as a parameter and the steric effect is modeled by the multiparametric segmental method. Use of μ^2 in place of μ gave poorer results. Attempts at improving the parameterization for hydrogen bonding and for charged substituents were unsuccessful. In parameterizing the hydrogen bonding it is necessary to take into account the levelling effect in aqueous phases which results in the formation of the greatest possible number of hydrogen bonds between water and substrate. The general form of the IMF equation may now be written as:

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + M\mu_X + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + I i_X + B_{DX} n_{DX} + B_{AX} n_{AX} + S\psi_X + B^{\circ}$$

Sets of log P values for XG where the hybridization of the C atom of the XC bond is the same throughout may be combined into a superset by using a parameter n_{CG} which accounts for the number of C atoms in G. Sets of log k' values determined under varying conditions of mobile phase or column packing may be combined by the ζ method. The IMF model provides an explanation for the success of the Hansch - Fujita model in the quantitative description of bioactivities. It also shows that bioactivities may be directly correlated with the IMF equation.

Our overall objective in this series of papers has been to determine the basis of the Hansch - Fujita model for the quantitative description of bioactivity. As the keystone of this model is the use of transport parameters such as the logarithms of the partition coefficient P and the high pressure liquid chromatography capacity factor k' it is therefore necessary to determine their nature. These and other related transport parameters are a measure of the difference in intermolecular forces between substrate - initial phase and those between substrate - final phase. Thus:

0097–6156/95/0606–0075$12.25/0

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

$$\tau = \Delta G_{imf} = G_{f,imf} - G_{i,imf} \qquad (1)$$

where G is the Gibbs free energy, i denotes the initial and f the final phase, and imf indicates intermolecular forces. These intermolecular forces are listed in Table 1 together with the quantities on which they depend. The data sets studied in this work are of the form **XG** where:

1. **X** is a variable substituent.
2. The skeletal group **G** to which **X** is bonded is usually constant.

In order to parameterize the intermolecular forces (imf) between **XG** and its surroundings it is necessary to consider:

1. The imf between **X** and its surroundings. These are modelled by parameters that describe the hydrogen bonding, bond moment, polarizability, and ionizability of **X**.
2. The effect of **X** on the imf between **G** and its surroundings. When **G** is sp^2 hybridized in whole or in part **X** may exert an effect on:
    a. Hydrogen bonding to **G** which may act as a hydrogen acceptor.
    b. Charge transfer involving **G** which may act as either an electron acceptor or an electron donor.
    c. Steric effects on the solvation of **G**.

When G is wholly sp^3 hybridized **X** may exert an effect on H atoms bound to the same C atom which affects the extent to which they form weak hydrogen bonds.

The effects of **X** on the imf between **G** and its surroundings are modelled by the electrical effect parameters. The steric effect of **X** on the solvation of **G** is accounted for by a steric parameterization.

Table 1. Intermolecular Forces and the Quantities Upon Which They Depend.

Intermolecular Force	**Quantity**
Molecule - molecule	
Hydrogen bonding (hb)	E_{hb}
Dipole - dipole (dd)	dipole moment
Dipole - induced dipole (di)	dipole moment, polarizability
Induced dipole - induced dipole (ii)	polarizability
Charge transfer (ct)	ionization potential, electron affinity
Ion - molecule	
ion - dipole (Id)	ionic charge, dipole moment
ion - induced dipole (Ii)	ionic charge, polarizability

In earlier work we have shown that transport parameters including log P, π, and chromatographic properties such as log k', R_M, and retention indices are quantitatively modelled by the intermolecular force (IMF) equation.[1-5] The IMF equation has been written in its most general form in the past as:

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + I i_X + B_{DX} n_{DX} + B_{AX} n_{AX} + S\psi_X + B^o \qquad (2)$$

where:

Q is the quantity to be correlated.

σ_{lX} is the **localized electrical effect** parameter. It is identical to the σ_I and σ_F constants.

σ_{dX} is the **intrinsic delocalized electrical effect** parameter.

σ_{eX} is the **electronic demand sensitivity electrical effect** parameter.

α is a **polarizability** parameter. It is defined by the equation:

$$\alpha \equiv (MR_X - MR_H)/100 = (MR_X - 0.0103)/100 \qquad (3)$$

where MR_X and MR_H are the group molar refractivities of X and H respectively.

$\mathbf{n_H}$ and $\mathbf{n_n}$ are **hydrogen bonding** parameters. n_H is equal to the number of OH or NH bonds in X while n_n is equal to the number of lone pairs on O or N atoms in X.

i is a parameter which accounts for the effects of the **charge on ionic groups**. It takes the value 1 when the side chain is ionized and 0 when it is not.

$\mathbf{n_D}$ and $\mathbf{n_A}$ are **charge transfer** parameters. n_D is 1 when X acts as an electron donor and 0 when it cannot. n_A is 1 when X can function as an electron acceptor and 0 when it cannot.

ψ is an appropriate **steric effect** parameterization. ψ may be monoparametric, using the υ steric parameters; multiparametric, using a branching model such as the simple branching equation, where:

$$S\psi = B_1 n_1 + B_2 n_2 + B_3 n_3 + \ldots = \sum_{i=1}^{m} B_i n_i \qquad (4)$$

or the segmental model, where:

$$S\psi = S_1 \upsilon_1 + S_2 \upsilon_2 + S_3 \upsilon_3 + \ldots = \sum_{i=1}^{m} S_i \upsilon_i \qquad (5)$$

or a composite model using a combination of the υ parameter and the branching model. In this work we consider several problems involving the improvement of the parameterization in the IMF equation.

1. Examination of the residuals in the correlation of log P or π for PhX with the IMF equation shows that the model is incomplete. This does not seem to be the case for AkX (Ak = alkyl) or for aromatic data sets in which a constant substituent results in a sufficiently large permanent dipole moment.
2. We have considered the possibility of improving the parameterization of hydrogen bonding by the use of the parameters proposed by Abraham and coworkers.[7]
3. We have also considered the possibility of improving the parameterization for ionic groups by means of a parameter which reflects acid and/or base strength.
4. We have reexamined the steric effect parameterization to determine whether it can be improved by use of the segmental method. We have also considered the possibility that the first segment may exert a steric effect only when it is larger than the half thickness of a benzene ring.

Method

We have correlated suitable test sets of transport parameters with appropriate variants of the IMF equation by means of multiple linear regression analysis. The substituent constants required for the correlations were taken when possible from our compilations. If values were unavailable they were estimated by methods described therein.[6,8-10] The dipole moments were taken from the compilations of McClellan.[11,12]

Test Set Requirements. The requirements for a suitable test set include:

1. A sufficient number of data points to permit a test of the IMF model. Our experience suggests that the minimum number is such that there will be at least three degrees of freedom per independent variable for chemical property data sets.
2. A wide range of substituent types in order to encompass a sufficient range of each of the variables and to minimize clustering.
3. Reliable determination of the transport parameters in the data set, preferably in the same laboratory in order to maximize the cancellation of experimental error.

The test sets chosen are reported in Table 3.

Composition of the Substituent Effect. We find it convenient in discussing our results to make use of the per cent contribution, C_i of each independent variable in the regression equation. This quantity is given by the expression:

$$C_i = 100|a_i x_i| / \sum_{i=1}^{m} |a_i x_i| \qquad (6)$$

where a_i is the regression coefficient of the i-th independent variable and x_i is its value for some reference group. In this work we have defined a hypothetical reference group for which:

$\sigma_l = \sigma_d = \mu = n_H = n_n = i = \upsilon_1 = \upsilon_2 = \upsilon_3 = n_D = n_A = 1$, $\sigma_e = 0.1$, $\alpha = 0.2$, $\zeta = 0.3$.

Results

PhX Transport Parameters. Our original parameterization of dipole-dipole and dipole-induced dipole interactions made use of electrical effect parameters to account for the dipole moment. We find that this is fully justified when:

1. The substituent is incapable of exerting a delocalized electrical effect.
2. The substrates have a large innate dipole moment.

In the case of PhX transport parameters neither of these requirements is met. It is therefore necessary to use μ_{PhX} as a parameter. The problem occurs because the dipole - dipole and dipole - induced dipole reactions depend on the magnitude of the moment and not its direction. The electrical effect constants will not represent the dipole moment unless the sign of the charge is taken into account if delocalized electrical effects can occur. Thus, correlation of dipole moments for PhX with the LDR equation:

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + h \qquad (7)$$

gives the regression equation 8 when the moments of groups (other than halogens) with negative σ_d values are assigned negative values and only symmetric groups are included in the data set.

$$\mu_{PhX} = 5.47(\pm 0.359)\sigma_{lX} + 4.30(\pm 0.446)\sigma_{dX} + 6.94(\pm 1.91)\sigma_{eX} + 0.420(\pm 0.172) \qquad (8)$$

$100R^2$, 96.23; F, 110.7; S_{est}, 0.342; S°, 0.222; Adj. $100R^2$, 95.69; P_D, 44.0 ± 5.25; η, 1.61 ± 0.411; n = 17. C_l, 52.3; C_d, 41.1; C_e, 6.63.

Inclusion of nonsymmetric substituents in the data set gave equation 9. The fit is poorer than that obtained for the symmetric groups alone. Of the regression coefficients only R has changed significantly. As the symmetric group data set does not include groups with a predominant donor delocalized effect such as OMe, SMe, and NH_2 **it is not a completely representative set** and some differences between equations 8 and 9 are therefore to be expected.

$$\mu_{PhX} = 4.98(\pm 0.448)\sigma_{lX} + 4.33(\pm 0.310)\sigma_{dX} + 1.50(\pm 1.47)\sigma_{eX} + 0.169(\pm 0.187) \qquad (9)$$

$100R^2$, 91.72; F, 147.8; S_{est}, 0.537; S°, 0.302; Adj. $100R^2$, 91.32; P_D, 46.5 ± 4.30; η, 0.347 ± 0.339; n = 44. C_l, 52.7; C_d, 45.8; C_e, 1.59.

The difference lies in the values of R and h which are not significant in equation 9. The values of L and D in equation 9 are not significantly different from those in equation 8. Correlation of dipole moments for symmetric MeX with the LDR equation gave the regression equation 10:

$$\mu_{MeX} = 5.11(\pm 0.497)\sigma_{lX} + 1.99(\pm 0.541)\sigma_{dX} + 0.0129(\pm 0.205) \qquad (10)$$

$100R^2$, 92.30; F, 59.91; S_{est}, 0.365; S°, 0.316; Adj. $100R^2$, 91.60; P_D, 28.0 ± 8.15; n = 13. C_l, 72.0; C_d, 28.0.

Inclusion of nonsymmetric groups gave the regression equation:

$$\mu_{MeX} = 6.33(\pm 0.574)\sigma_{lX} + 2.00(\pm 0.334)\sigma_{dX} - 0.0803(\pm 0.235) \qquad (11)$$

$100R^2$, 83.81; F, 95.75; S_{est}, 0.593; S°, 0.418; Adj. $100R^2$, 83.38; P_D, 24.0 ± 4.44; n = 40. C_l, 76.0; C_d, 24.0.

The values of μ for X = NH_2, NMe_2, and PH_2 were assigned negative values based on the results obtained for the symmetric groups in equation 10. The agreement between the coefficients of equations 10 and 11 is fairly good. Clearly, when the data set includes amino, alkylamino, and similar groups bonded to sp^3 hybridized carbon electrical effect parameters cannot serve to represent the effect of the dipole moment. The dipole moments used in these correlations are reported in Table 2.

Table 2. Dipole moment data sets correlated with the LDR equation.

μ, PhX. X, μ; tBu, (-) 0.52; Me, (-) 0.37; H, 0; F, 1.66; Cl, 1.70; Br, 1.70; I, 1.71; Ph, 0; C$_2$H, 0.71; CF$_3$, 2.86; CCl$_3$, 2.14; CN, 4.14; NO$_2$, 4.28; SiMe$_3$, 0.42; SF$_5$, 3.44; C$_2$Ph, 0; C$_2$CF$_3$, 3.38; *OH,(-) 1.40; NH$_2$,(-) 1.49; SH, (-) 1.21; PH$_2$, (-) 1.11; SeH, (-) 1.10; NHMe, (-) 1.77; NMe$_2$, (-) 1.60; OMe (-) 1.36; PMe$_2$, (-) 1.31; SMe, (-) 1.29; N$_3$, 1.56; Et, (-) 0.37; Pr, (-) 0.37; iPr, (-) 0.40; Bu, (-) 0.37; cPr, (-) 0.48; CH=CH$_2$, (-) 0.13; OEt, (-) 1.38; CO$_2$Me, 1.92; CO$_2$Et, 1.849; CO$_2$iPr, 1.97; CO$_2$H, 1.86; COCl, 3.18; Ac, 2.88; CHO, 2.92; Bz, 2.96.*

μ,MeX. X, μ; H, 0; Me, 0; F, 1.8572; Cl, 1.869; Br, 1.797; I, 1.618; CF$_3$,2.321; CCl$_3$,1.755; Ph,0.37; C$_2$H,0.7809; CN,3.9185; NO$_2$, 3.57; C$_6$Cl$_5$, 1.56; *N$_3$, 2.17; SH, 1.52; SCl, 2.00; NH$_2$, (-) 1.296; OH, 1.77; COCl, 2.71; OMe, 1.31; SSMe, 1.985; CO$_2$Me, 1.706; Ac, 2.93; COF, 2.96; SCN, 3.34; SeMe, 1.41; cPr, 0.139; NMe$_2$, (-) 0.612; CHO,2.69; OPh,1.36; SPh,1.29; PH$_2$,(-)1.100; SeCN,3.91; CO$_2$Et, 1.84; CH=CH$_2$, 0.364; SO$_2$Ph, 4.73; Bz, 2.96.*

Groups in italics are nonsymmetric, other groups are symmetric.

Table 3. Test Data Sets.

1. Log P, 1-octanol/water, PhX.[a] *X, Log P.* H, 2.13; Me, 2.69; Et, 3.15; iPr, 3.66; tBu, 4.11; CH=CH$_2$, 2.95; Ph, 4.04; CH$_2$CH=CH$_2$, 3.23; CH$_2$Ph, 4.14; cPr, 3.27; F, 2.27; Cl, 2.84; Br, 2.99; I, 3.25; CF$_3$, 2.79; CCl$_3$, 2.92[x]; SMe, 2.74; SPh, 4.45; SeMe, 2.87; SiMe$_3$, 4.72; SCF$_3$, 3.57; SSPh, 4.41; PMe$_2$, 2.57; PhCH$_2$CH$_2$, 4.82; 2-Thienyl, 3.74; CH$_2$Br, 2.92; CH$_2$Cl, 2.30; SF$_5$, 3.76; OMe, 2.11; OEt, 2.51; OPh, 4.21; CHO, 1.48; Ac, 1.58; Bz, 3.18; CO$_2$Me, 2.12; CO$_2$Et, 2.64; CO$_2$iPr, 3.18; OAc, 1.49; CN, 1.56; NMe$_2$, 2.31; SCN, 2.54; OCF$_3$, 3.17; CH$_2$CN, 1.56; CH$_2$OMe, 1.35; CH$_2$OAc, 1.96; CH$_2$Ac, 1.44; CH$_2$CO$_2$Et, 2.28; NO$_2$, 1.85; SOMe, 0.55[x]; SOPh, 2.06[x]; SO$_2$Me, 0.50[x]; SO$_2$CF$_3$, 2.68[x]; SO$_2$Ph, 2.40; PO(OMe)$_2$, 0.95[x]; NH$_2$, 0.90; NHMe, 1.66; NHAc, 1.16; CONH$_2$, 0.64; SO$_2$NH$_2$, 0.31; OH, 1.46; CH$_2$OH, 1.10; NHSO$_2$Me, 0.95; CO$_2$H, 1.87; CH$_2$CO$_2$H, 1.41; CH$_2$NH$_2$, 1.09; CH$_2$CH$_2$NH$_2$, 1.41; CH$_2$CH$_2$CO$_2$H, 1.84; CH=CHCO$_2$H, 2.13; NHNH$_2$, 1.25; NHCONH$_2$, 0.83; OSO$_2$Me, 1.25; COEt, 2.19; POPH$_2$, 2.87; N=NPh, 3.82; N$_3$, 2.59; OBz, 3.59; NHBz, 2.70; CH=CHNO$_2$, 2.24[x]; CH$_2$CHO, 1.96; Pr, 3.57; Bu, 4.26; NHEt, 2.26; NHPh, 3.50; NEt$_2$, 3.31; NPh$_2$, 5.74; CH=C(CN)$_2$, 2.12; CH=CHPh, 4.81; 4'-HOC$_6$H$_4$, 3.20; CH$_2$SiMe$_3$, 4.13; NHCO$_2$Me, 1.76.

2. Log P, propylene glycol dipelargonate/water, PhX.[b] *X, Log P.* H, 2.36; Me, 2.89; Et, 3.37; CH=CH$_2$, 3.03; CH$_2$CH=CH$_2$, 3.65; CF$_3$, 3.26; F, 2.50; Cl, 3.08; Br, 3.27; I, 3.48; CN, 1.66; NO$_2$, 2.16; NH$_2$, 0.95; NHMe, 1.17; NMe$_2$, 2.52; OH, 1.16; OMe, 2.41; OAc, 1.57; CHO, 1.57; Ac, 1.63; Bz, 3.40; CO$_2$H, 1.15; CONH$_2$, -0.36; NHAc, 0.40; NHCONH$_2$, -0.55; NHCO$_2$Me, 1.84; SOMe, -0.41; SO$_2$Me, 0.47; SO$_2$NH$_2$, -0.03; NHSO$_2$Me, 0.77; CO$_2$Me, 2.32; CO$_2$Et, 2.84; CO$_2$iPr, 3.30; CH$_2$OH, 0.61; CH$_2$NH$_2$, 0.18; CH$_2$CO$_2$H, 0.65; CH$_2$CH$_2$NH$_2$, 0.52.

3. Log P, 1-octanol/water, MeX.[a] *X, Log P.* H, 0; Me, 1.81; Et, 2.36; Pr, 2.89, iPr, 2.76; Bu, 3.23; tBu, 3.11; CH=CH$_2$, CH$_2$CH=CH$_2$, 2.40; Ph, 2.73; CH$_2$Ph, 3.15; CCl$_3$, 2.49; F, 0.51; Cl, 0.91; Br, 1.19; I, 1.51; SPh, 2.74; CH$_2$Cl, 1.43; CH$_2$Br, 1.61; CH$_2$I, 2.00; E-CH=CHMe, 2.31; SSMe, 1.77; NMe$_2$, 0.16; CN, -0.34; Ac, -0.24; NO$_2$, -0.35; OMe, 0.10; OPh, 2.11; CO$_2$Me, 0.18; CO$_2$Et, 0.66; CH$_2$CN, 0.16; CH$_2$Ac, 0.26; OAc, 0.18; C$_2$Me, 1.46; SO$_2$Ph, 0.47; OH, -0.77; NH$_2$, -0.57; C$_2$H, 0.94; NHAc, -1.05; CH$_2$OH, -0.31; NHPh, 1.66; CO$_2$H, -0.31; CH$_2$CO$_2$H, 0.25; CH$_2$NH$_2$, -0.13; CONH$_2$, -0.66; NHEt, 0.15; CH$_2$C$_2$H, 1.46; Me$_3$Si, 3.85; MeSO, -1.85.

Table 3. *Continued*

4. Log P, 1-octanol/water, EtX.[a] *X, Log P.* H, 1.81; Me, 2.36; Et, 2.89; Pr, 3.39; CH=CH_2, 2.40; tBu, 3.82; Ph, 3.15; $PhCH_2$, 3.68; Cl, 1.43; Br, 1.61; I, 2.00; EtS, 1.95; PhS, 3.20; CH_2Cl, 2.04; CH_2Br, 2.10; Me_2N, 0.70; NC, 0.16; Ac, 0.26; O_2N, 0.18; HO_2CCH_2, 0.79; H_2NCH_2Et, 0.48; EtO, 0.89; PhO, 2.51; EtO_2C, 1.21; $AcCH_2$, 0.91; AcO, 0.66; Bz, 2.19; BzO, 2.64; Et_2N, 1.44; HO, -0.31; H_2N, -0.13; $H_2NHCONH$, -0.74; H_2NCO, -0.15; MeNH, 0.15; HC_2, 1.46; $HOCH_2$, 0.25; HO_2C. 0.25; PhNH, 2.26;

11. Log P, 1-octanol/water, $Y_{Ak}X$.[a] Includes set 3 and 4 +: *XY, Log P.* iPrH, 2.36; PrH, 2.36; BuH, 2.89; tBuH, 2.76; Pe, 3.39; iPrMe, 2.76; tBuMe, 3.11; BuMe, 3.39; EtPr, 3.39; $(iPr)_2$, 3.85; H_2C=$CHCH_2Me$, 2.40; PhiPr, 3.66; PhBu, 4.11; PhtBu, 4.11; BrPr, 2.10; Me_2NBu, 1.70; AcPr, 0.91; AcBu, 1.38; HCOPe, 1.78; O_2NPr, 0.87; O_2NBu, 1.47; O_2NPe, 2.01; EtOBu, 2.03; $AcCH_2Pr$, 1.38; BzOiPr, 3.18; HOPr, 0.25; HOiPr, 0.05; HOBu, 0.88; HOtBu, 0.35; HOPe, 1.56; HOHx, 2.03; HOHp, 2.57; HODd, 5.13; H_2NiPr, -0.03; H_2NBu, 0.88; H_2tBu, 0.40; H_2NPe, 1.49; H_2NHx, 2.06; $H_2NCOiPr$, -0.21; HC_2Pr, 1.98; $HOCH_2Pr$, 0.88; $HOCH_2iPr$, 0.76; $HOCH_2Bu$, 1.56; $HOCH_2Pe$, 2.03; HO_2CPr, 0.79; HO_2CPe,1.88; H_2NCH_2Pr, 0.88; H_2NCH_2iPr, 0.73; H_2NCH_2Bu, 1.49; H_2NCH_2Pe, 2.06; CH_2C_2H, 1.98; Me_3SiPr,3.84; Me_3SiBu, 4.20; Me_3SiOc, 4.24; HSBu, 2.28.

12. Log P, 1-octanol/water, $Y_{Cx}X$.[a] Includes set 1 +: *XY, Log P.* H_2C=CH_2, 1.13; NCCH=CH_2, -0.92; MeCH=CH_2, 1.77; $HOCH_2CH$=CH_2, 0.17; H_2NCH_2CH=CH_2, 0.03; EtCH=CH_2, 2.40; EtOCH=CH_2, 1.04; $(H_2C$=$CH)_2$, 1.99; BzCH=CH_2, 1.88; $PhOCH_2CH$=CH_2, 2.94; PhCH=CH_2, 2.95; $(H_2C$=$CH)_2CH_2$, 1.48; 2-HNh, 3.37; 2-MeNh, 3.86; 2-HONh, 2.84; 2-H_2NNh, 2.28; 1-HNh, 3.37; 1-MeNh, 3.87; 1-HONh, 2.84; 1-AcONh, 2.78; 4-HOC_6H_4Ph, 3.20; HAnc, 4.45; HPnt, 4.46; HO_2CCH=CHPh, 2.13; MeCH=CHPh, 3.35; O_2NCH=CHPh, 2.24; $HOCH_2CH$=CHPh, 1.95; NCCH=CHPh, 1.96; HCOCH=CHPh, 1.90; MeO_2CCH=CHPh, 2.62; AcCH=CHPh, 2.07; EtO_2CCH=CHPh, 2.99; PhCH=CHPh, 4.25; BzCH=CHPh, 3.08; H_2C=CHPh, 2.95.

21. Log P, 1-octanol/water, $X(CH_2)_nPh$.[a] *X, n, Log P, ...* H, 1, 2.73, 2, 3.15, 3, 3.68, 4, 4.26; Me, 1, 3.15, 2, 3.68, 3, 4.26; Et, 1, 3.68; Pr, 1, 4.26; F, 3, 2.95; Cl, 1, 2.30, 2, 2.95, 3, 3.55; Br, 1, 2.92, 2, 3.09, 3, 3.72; I, 3, 3.90; OH, 1, 1.10, 2, 1.36; OMe, 1, 1.35, 3, 2.70; OAc, 1, 1.96, 2, 2.30, 3, 2.77; NH_2, 1, 1.09, 2, 1.41, 3, 1.83; NMe_2, 1, 1.98, 3, 2.73; NO_2, 1, 1.75, 2, 2.08; CO_2H, 1, 1.41, 2, 1.84, 3, 2.42; CO_2Me, 1, 1.83, 2, 2.32, 3, 2.77; CO_2Et, 1, 2.28, 2, 2.73; CHO, 1, 1.78; Ac, 1, 1.44, 3, 2.42; $CONH_2$, 1, 0.45, 2, 0.91, 3, 1.41; CN, 1, 1.56, 2, 1.72, 3, 2.21; SCN, 1, 1.99[x]; SO_2NH_2, 2, 0.85[x]; CH=CH_2, 1, 3.23; $SiMe_3$, 1, 4.13; OPh, 1, 3.79; Ph, 1, 4.14, 2, 4.82; OBz, 1, 3.97; OCONHMe, 2, 1.93[x]; CH=NOH, 2, 1.56[x]; CH_2Cl, 1, 2.95, 2, 3.55; CH_2Br, 1, 3.09, 2, 3.72; CH_2F, 2, 2.95[x]; CH_2I, 2, 3.90; CH_2OH, 1, 1.36; CH_2OAc, 1, 2.30, 2, 2.77; CH_2NH_2, 1, 1.41, 2, 1.83; CH_2CO_2H, 1, 1.84, 2, 2.42; CH_2CO_2Me, 1, 2.32[x], 2, 2.77[x]; CH_2CN, 1, 1.72, 2, 2.21; CH_2Ph, 1, 4.82; CH_2NMe_2, 2, 2.73; CH_2NO_2, 1, 2.08; CH_2Ac, 2, 2.42; CH_2CO_2Et, 1, 2.73.

31. τ_W, 4-$XPnSO_2NH_2$.[c] *X,* τ_{MeOH}, τ_{MeCN}. H, 0, 0; Me, 0.57, 0.59; Cl, 0.67, 0.80; Br, 0.83, 0.97; I, 1.09, 1.03; iPr, 1.45, 1.09; Ph, 2.01, 1.50; CN, 0.21, 0.40; Ac, 0.50; 0.46; OMe, 0.46; 0.34; NO_2, 0.31, 0.51; OBu, 1.94, 1.47; OHx, 2.99; 2.27; SO_2NH_2, -0.50, -0.22; $CONH_2$, -0.29, -0.39; NH_2, -0.58, -0.59; $NHSO_2Me$, 0.05, 0.01; NHAc, 0.29, 0.24; OH, -0.33, -0.22.

Continued on next page

Table 3. *Continued*

32m. Log k', 3-XPnNHCOCH$_2$Ph.[d] *X, Log k'(50% v/v aq. MeOH, 70% v/v aq. MeOH, 40:5:55 v/vMeOH-THF-H$_2$O).* H, 0.91, 0.11, 0.97; Me, 1.22, 0.31, 1.23; Et, 1.48, 0.49, 1.54; F, 1.11, 0.23, 1.25; Cl, 1.40, 0.44, 1.53; Br, 1.50, 0.50, 1.63; I, 1.60, 0.57, 1.74; CF_3, 1.62, 0.51, 1.79; OH, 0.46, -0.28, 0.60; NH_2, 0.34, -0.34, 0.39; NHMe, 0.71, -0.08, 0.75; NHEt, 0.98, 0.11, 1.00, OMe, 1.00, 0.15, 1.04; CN, 0.93, 0.06, 1.02; NO_2,1.12, 0.23, 1.29; Ac, 0.90, 0.04, 0.87; OAc, 0.90, 0.01, 0.89; OEt, 1.28, 0.33, 1.29; OSO_2Me, 0.76, -0.14, 0.87.

32p. Log k', 4-XPnNHCOCH$_2$Ph.[d] *X, Log k'(50% v/v aq. MeOH, 70% v/v aq. MeOH, 40:5:55 v/vMeOH-THF-H$_2$O).* H, 0.91, 0.11, 0.97; Me, 1.20, 0.29, 1.22; F, 1.01, 0.16, 1.11; Cl, 1.38, 0.43, 1.52; Br, 1.50, 0.50, 1.61; OH, 0.33, -0.39, 0.43; OMe, 0.89, 0.06, 0.90; CN, 0.90, 0.06, 0.86; NO_2, 1.17, 0.28, 1.38, Ac, 0.94, 0.03, 1.03.

33. Log k', PhX.[d] *X, Log k'(50% v/v aq. MeOH, 70% v/v aq. MeOH, 40:5:55 v/vMeOH-THF-H$_2$O)*; H, .96, .36, 1.03; Me, 1.30, 0.61, 1.38; Et, 1.62, 0.81, 1.71; F, 0.98, 0.33, 1.07; Cl, 1.32, 0.58, 1.42; Br, 1.42, 0.65, 1.51; I, 1.57, 0.76, 1.69; CF_3, 1.43, 0.55, 1.56; OH, 0.17, -0.32, 0.33; NH_2, 0.13, -0.32, 0.16; NHMe, 0.64, 0.06, 0.65; NHEt, 0.87, 0.25, 0.92; OMe, 0.94, 0.33, 0.98; CN, 0.54, -0.07, 0.50; NO_2, 0.75, 0.15, 0.82; Ac, 0.64, 0.01, 0.59.

34. Log k', XPh.[e] *X, Log k'(aq. MeOH, aq. MeCN, aq. THF)*;. H, 1.430, 1.321, 1.925; F, 1.647, 1.421, 1.837; Cl, 2.356, 2.035, 2.511; Br, 2.510, 2.183, 2.575; I, 2.795, 2.414, 2.851; Me, 2.166, 1.877, 2.358; iPr, 3.124, 2.818, 3.047; Ph, 3.633, 3.140, 3.341; NO_2, 1.486, 1.331, 1.649; CN, 1.301, 0.981, 1.103; CHO, 1.134, 0.866, 0.928; OMe, 1.767, 1.423, 1.715; OEt, 2.146, 1.818, 2.161; Ac, 1.540, 1.048, 0.965; Bz, 2.922, 2.374, 2.389; CO_2Me, 1.966, 1.448, 1.600; CO_2Et, 2.450, 1.930, 1.922; CO_2Pr, 2.961, 2.410, 2.544; CO_2H, 1.051, 0.650, 1.178; OH, 0.613, 0.426, 1.051.

35. k', XPh.[f] *X, k'(mobile phase 60 : 40 MeOH - H$_2$O).(PFP, HFIPP, HFD, C$_{10}$);* H, 0.26, 0.48, 0.99, 2.14; Me, 0.36, 0.66, 1.53, 3.94; Et, 0.50, 0.91, 2.29, 6.77; Pr, 0.70, 1.29, 3.67, 12.53; Bu, 0.98, 1.82, 5.74, 23.06; tBu, 0.80, 1.58, 4.84, 15.82; F, 0.32, 0.54, 1.22, 2.20; Cl, 0.38, 0.61, 1.36, 3.91; Br, 0.42, 0.61, 1.32, 4.59; I, 0.46, 0.60, 1.29, 5.98, CH_2Br, 0.35, 0.50, 0.87, 1.60; OH, 0.12, 0.14, 0.25, 0.53; CH_2OH, 0.13, 0.18, 0.26, 0.51; CN, 0.23, 0.35, 0.59, 0.83; CH_2CN, 0.20, 0.28, 0.45, 0.78; NO_2, 0.26, 0.41, 0.77, 1.47; OMe, 0.31, 0.43, 0.79, 2.03; NH_2, 0.12, 0.20, 0.25, 0.41; CHO, 0.21, 0.29, 0.51, 0.84; Ac, 0.25, 0.36, 0.65, 0.98; Bz, 0.45, 0.83, 1.47, 4.35; Ph, 0.76, 0.93, 2.09, 5.53; CO_2H, --, 0.03, 0.05, 0.25.

a. Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979. b. Leahy, D. E., Taylor, P. J., and Wait, A. R., *Quant. Struct.-Act. Relat.*, **8**, 17-31 (1989). c. Altomare, C., Carotti, A., Cellamare, S., and Ferappi, M., *Int. J. Pharmaceutics* **56**, 273-281 (1989). d. Yamagami, C., Takami, H., Yamamoto, K., Miyoshi, K., and Takao, N., *Chem. Pharm. Bull.* **32**, 4994-5002 (1984). e. Miyake, K., Mizuno, N., and Terada, H., *Chem. Pharm. Bull.* **34**, 4787-4796 (1986). Log k'$_0$ are obtained by extrapolation to 0 % organic solvent in the mobile phase. f. Sadek, P. C., and Carr, P. W., *J. Chromatogr.* **288**, 25-41 (1984). x. Excluded from the best regression equation.

Abbreviations: Pe, pentyl; Hx, hexyl, Hp, heptyl; Oc, octyl; Dd, dodecyl; Nh, naphthyl; Anc, anthryl; Pnt, phenanthryl; Pn, phenylene; PFP, pentafluorophenyldimethylsilane; HFIPP, heptafluoroisopropoxypropyldimethylsilane; HPD, heptadecafluorodecyldimethylsilane; C$_{10}$, decyldimethylsilane.

Dipole Moment Parameterization in the IMF Equation. Transport parameter sets 1 and 2 (Table 3) were correlated with the IMF equation. The correlation equation used had the form:

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + I i_X + M\mu_X + S_1 \upsilon_{1X} + S_2 \upsilon_{2X} + B^o \qquad (12)$$

The best regression equations obtained were for set 1:

$$\begin{aligned} LogP_{PhX} = {} & 1.96(\pm 0.322)\sigma_{lX} + 2.39(\pm 0.732)\sigma_{eX} + 6.67(\pm 0.480)\alpha_X - 0.368(\pm 0.0687)n_{HX} \\ & - 0.172(\pm 0.0309)n_{nX} - 0.452(\pm 0.0491)\mu_X + 0.726(\pm 0.179)\upsilon_{1X} \\ & + 0.363(\pm 0.191)\upsilon_{2X} + 1.99(\pm 0.157) \qquad (13) \end{aligned}$$

100R^2, 91.60; Adj. 100R^2, 90.78; F, 96.77; S$_{est}$, 0.340; S^o, 0.308; n, 80.
C$_l$, 34.9; C$_e$, 4.26; C$_\alpha$, 23.8; C$_\mu$, 8.06; C$_{\upsilon 1}$, 12.9; C$_{\upsilon 2}$, 6.47; C$_{nH}$, 6.56; C$_{nn}$, 3.06.

and for set 2:

$$\begin{aligned} LogP_{PhX} = {} & 1.79(\pm 0.759)\sigma_{lX} + 6.38(\pm 1.59)\alpha_X - 0.751(\pm 0.136)n_{HX} - 0.187(\pm 0.0641)n_{nX} \\ & - 0.505(\pm 0.278)i_X - 0.408(\pm 0.120)\mu_X + 0.862(\pm 0.449)\upsilon_{1X} + 2.08(\pm 0.265) \qquad (14) \end{aligned}$$

100R^2, 90.44; Adj. 100R^2, 88.31; F, 35.12; S$_{est}$, 0.431; S^o, 0.354; n, 34.
C$_l$, 29.0; C$_\alpha$, 20.6; C$_\mu$, 13.2; C$_{\upsilon 1}$, 13.9; C$_{nH}$, 12.1; C$_{nn}$, 3.02; C$_i$, 8.16.

Correlating set 1 with a form of equation 12 in which μ is replaced by μ^2 gives the regression equation:

$$\begin{aligned} LogP_{PhX} = {} & 1.05(\pm 0.334)\sigma_{lX} + 2.67(\pm 0.834)\sigma_{eX} + 7.42(\pm 0.524)\alpha_X - 0.374(\pm 0.0743)n_{HX} \\ & - 0.203(\pm 0.0347)n_{nX} - 0.0587(\pm 0.00964)\mu_X^2 + 0.587(\pm 0.206)\upsilon_{1X} + 1.89(\pm 0.168) \qquad (15) \end{aligned}$$

100R^2, 88.91; Adj. 100R^2, 88.00; F, 82.48; S$_{est}$, 0.392; S^o, 0.351; n, 80.
C$_l$, 25.0; C$_e$, 6.36; C$_\alpha$, 35.3; C$_{\mu 2}$, 5.60; C$_{\upsilon 1}$, 14.0; C$_{nH}$, 8.90; C$_{nn}$, 4.82.

A comparison of μ and μ^2 as parameters shows that best results are obtained with the former.

Steric Effect Parameterization in the IMF Equation. Replacement of the monoparametric υ parameterization of the steric effect by the multiparametric segmental model further improved the description. We have also considered the replacement of the υ_1 parameter with the υ_Δ parameter defined as υ_1 minus the υ_{mn} value for Ph (the minimal value of its steric parameter). This is equivalent to the difference between the van der Waals radius of segment 1 and the half thickness of the benzene ring. Correlation of set 1 with this variant of equation 12 gave as the best regression equation:

$$\begin{aligned} LogP_{PhX} = {} & 1.84(\pm 0.310)\sigma_{lX} + 2.49(\pm 0.706)\sigma_{eX} + 6.73(\pm 0.468)\alpha_X \\ & - 0.375(\pm 0.0667)n_{HX} - 0.182(\pm 0.0293)n_{nX} - 0.440(\pm 0.0466)\mu_X \\ & + 1.03(\pm 0.223)\upsilon_{\Delta X} + 0.387(\pm 0.193)\upsilon_{2X} + 2.33(\pm 0.124) \qquad (16) \end{aligned}$$

100R^2, 92.04; Adj. 100R^2, 91.27; F, 102.7; S_{est}, 0.331; S°, 0.299; n, 80.
C_l, 31.4; C_e, 4.26; C_α, 23.0; C_μ, 7.52; $C_{\upsilon\Delta 1}$, 17.7; $C_{\upsilon 2}$, 6.62; C_{nH}, 6.40; C_{nn}, 3.10.

The results obtained with the υ_Δ parameter are very slightly better than those obtained with the υ_1 parameter. The difference is probably not significant.

ZSO and probably Z^1Z^2PO groups as well show large deviations from the model. Other groups which deviate significantly are SO_2Me, $NHCONH_2$, $NHCO_2Me$, and $CH{=}CHNO_2$. Data points excluded from the best regression equation are indicated in Table 3 by the superscript x.

Hydrogen Bond Parameterization. in the IMF Equation. Correlation of set 1 with a variant of equation 12 in which the n_H parameter is replaced by the a_H parameter derived from the values reported by Abraham et al. gave the best regression equation:

$$LogP_{PhX} = 2.60(\pm 0.329)\sigma_{lX} + 0.500(\pm 0.213)\sigma_{dX} + 2.25(\pm 0.881)\sigma_{eX} + 6.73(\pm 0.517)\alpha_X - 0.198(\pm 0.0348)n_{nX} - 0.478(\pm 0.172)i_X - 0.562(\pm 0.0543)\mu_X + 0.703(\pm 0.205)\upsilon_{1X} + 0.633(\pm 0.200)\upsilon_{2X} + 1.96(\pm 0.174) \qquad (17)$$

100R^2, 90.53; Adj. 100R^2, 89.46; F, 74.36; S_{est}, 0.364; S°, 0.329; n, 80.
C_l, 35.9; C_d, 6.90; C_e, 3.10; C_α, 18.6; C_μ, 7.77; $C_{\upsilon 1}$, 9.71; $C_{\upsilon 2}$, 8.74; C_{nn}, 2.74; C_i, 6.60

Not only were the results inferior to those obtained with our previous parameterization, a_H was not a significant parameter.

Ionic Group Parameterization.in the IMF Equation. We have examined the use of the i° parameter, defined by the relationship:

$$i^o = (14 - pK)/10 \qquad (18)$$

where pK is pK_a for an acidic and pK_b for a basic substituent. When used as a variable in place of i in equation 12 it gave results for set 1 which were inconclusive as i° was not a significant parameter. For set 2 the best regression equation was:

$$LogP_{PhX} = 4.26(\pm 0.884)\sigma_{lX} + 9.91(\pm 2.10)\alpha_X - 0.735(\pm 0.132)\mu_X - 0.259(\pm 0.0883)n_{nX} - 1.03(\pm 0.355)i^o + 2.02(\pm 0.302) \qquad (19)$$

100R^2, 78.30; Adj. 100R^2, 75.31; F, 20.20; S_{est}, 0.625; S°, 0.513; n, 34.
C_l, 47.3; C_α, 22.0; C_μ, 16.3; C_{nn}, 2.88; C_{io}, 11.5.

Our original parameterization for ionic groups seems much better than the i° parameter. We have also correlated log P values for MeX (set 3) and EtX (set 4) with a variant of equation 12 having an additional term in υ_3. The best regression equations obtained are:

$$\log P_{MeX} = 0.483(\pm 0.216)\sigma_{dX} + 5.05(\pm 1.01)\sigma_{eX} - 0.484(\pm 0.0456)\mu_X + 8.95(\pm 0.615)\alpha_X - 0.425(\pm 0.0976)n_{HX} - 0.298(\pm 0.0371)n_{nX} + 1.54(\pm 0.112) \qquad (20)$$

100R^2, 94.55; Adj. 100R^2, 93.87; F, 112.7; S_{est}, 0.302; S°, 0.254; n, 46.
C_d, 12.1; C_e, 12.7; C_α, 44.9; C_μ, 12.2; C_{nH}, 10.7; C_{nn}, 7.48.

for set 3 and

$$\log P_{EtX} = 0.692(\pm 0.294)\sigma_{dX} + 5.82(\pm 1.53)\sigma_{eX} - 0.444(\pm 0.0777)\mu_X + 9.26(\pm 0.874)\alpha_X - 0.273(\pm 0.0545)n_{HX} - 0.463(\pm 0.223)n_{nX} + 1.92(\pm 0.175) \quad (21)$$

100R^2, 92.67; Adj. 100R^2, 91.31; F, 54.78; S_{est}, 0.373; S°, 0.305; n, 33.
C_d, 16.1; C_e, 13.5; C_α, 43.0; C_μ, 10.3; C_{nH}, 6.34; C_{nn}, 10.7.

for set 4:. These results support those obtained for the aromatic data sets.

Combination of Data Sets for Skeletal Groups with the Same Hybridization. Log P values for XG where the skeletal group G contains carbon atoms in the same hybridization state were combined into a single data set and correlated with the IMF equation in the form:

$$Q_{XG} = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + M\mu_X + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + S_1\upsilon_{1X} + S_2\upsilon_{2X} + S_3\upsilon_{3X} + B_{CG}n_{CG} + B^o \quad (22)$$

With groups G hybridized sp^2 (G is phenyl, vinyl, styryl, 1- or 2- naphthyl, 4-biphenyl, anthryl, or phenanthryl) (set 12, Table 3) the best regression equation is:

$$P_{XG} = 1.64(\pm 0.280)\sigma_{lX} + 2.00(\pm 0.692)\sigma_{eX} - 0.421(\pm 0.0428)\mu_X + 6.64(\pm 0.415)\alpha_X - 0.397(\pm 0.059)n_{HX} - 0.186(\pm 0.028)n_{nX} + 0.690(\pm 0.157)\upsilon_{1X} + 0.281(\pm 0.017)n_{CG} + 0.483(\pm 0.158) \quad (23)$$

100R^2, 90.92; Adj. 100R^2, 90.32; F, 131.4; S_{est}, 0.344; S°, 0.314; n, 114.
C_l, 31.9; C_e, 3.90; C_α, 25.8; C_μ, 8.19; $C_{\upsilon 1}$, 13.4; C_{nH}, 7.72; C_{nn}, 3.62; C_{nC}, 5.46.

Substitution of υ_Δ for υ_1 gives as the best regression equation:

$$\log P_{XG} = 1.55(\pm 0.279)\sigma_{lX} + 2.03(\pm 0.692)\sigma_{eX} - 0.394(\pm 0.0415)\mu_X + 6.95(\pm 0.405)\alpha_X - 0.381(\pm 0.059)n_{HX} - 0.194(\pm 0.027)n_{nX} + 0.943(\pm 0.216)\upsilon_{\Delta X} + 0.274(\pm 0.016)n_{CG} + 0.765(\pm 0.138) \quad (24)$$

100R^2, 90.91; Adj. 100R^2, 90.31; F, 131.3; S_{est}, 0.345; S°, 0.314; n, 114.
C_l, 31.9; C_e, 4.19; C_α, 28.6; C_μ, 8.11; $C_{\upsilon\Delta}$, 9.71; C_{nH}, 7.85; C_{nn}, 3.99; C_{nC}, 5.64.

There is no real difference between the results obtained using υ_1 and those obtained using υ_Δ.
When G is alkyl (has only sp^3 hybridized C atoms) (set 11, Table 3) the best regression equation is:

$$\log P_{XG} = 1.18(\pm 0.270)\sigma_{lX} + 0.642(\pm 0.132)\sigma_{dX} + 4.79(\pm 0.637)\sigma_{eX} - 0.588(\pm 0.0450)\mu_X + 7.91(\pm 0.456)\alpha_X - 0.395(\pm 0.0521)n_{HX} - 0.276(\pm 0.0262)n_{nX} + 0.458(\pm 0.168)\upsilon_{2X} + 0.522(\pm 0.0188)n_{CG} + 0.804(\pm 0.0953) \quad (25)$$

$100R^2$, 94.09; Adj. $100R^2$, 93.71; F, 219.3; S_{est}, 0.309; S^o, 0.253; n, 134.
C_l, 19.3; C_d, 10.5; C_e, 7.82; C_α, 25.8; C_μ, 9.61; $C_{\upsilon 2}$, 7.47; C_{nH}, 6.46; C_{nn}, 4.51; C_{nC}, 8.52.

We have examined a second example in which X is bonded to sp^3 hybridized carbon atoms in G where G is $(CH_2)_nPh$ (set 21, Table 3). The correlation equation used is again eq. 22 but in this case the n_{CG} parameter is equal to the number of CH_2 groups in G, the Ph group is not considered as it is constant throughout the data set. The best regression equation obtained is:

$$\begin{aligned}\log P_{XG} = {} & 1.15(\pm 0.320)\sigma_{lX} + 0.726(\pm 0.168)\sigma_{dX} + 4.34(\pm 0.726)\sigma_{eX} - 0.521(\pm 0.054)\mu_X \\ & + 8.35(\pm 0.575)\alpha_X - 0.310(\pm 0.06)n_{HX} - 0.255(\pm 0.024)n_n \\ & + 0.420(\pm 0.042)n_{CG} + 2.46(\pm 0.127) \qquad (26)\end{aligned}$$

$100R^2$, 93.61; Adj. $100R^2$, 92.90; F, 113.5; S_{est}, 0.272; S^o, 0.271; n, 71.
C_l, 21.0; C_d, 13.2; C_e, 7.91; C_α, 30.4; C_μ, 9.49; C_{nH}, 5.65; C_{nn}, 4.64; C_{nC}, 7.66.

Except for a lack of dependence on υ_2 the resemblance between the coefficients of equation 25 and those of equation 26 is striking.

Combination of Data Sets Studied in Different Media. We have shown that it is possible to combine data sets studied under different conditions or having different structural features into a single data set by means of internal parameters. This technique, which we have called the Zeta method[14], requires the definition of the experimentally determined quantities for some reference substituent X^o common to all of the data sets to be combined as values of the parameter ζ. The parameter ζ then accounts for the conditions or structural features which differentiate between the subsets. The reference substituent X^o must be chosen so that it is sensitive to the conditions of interest. In all of the sets we have studied we have chosen the OH group as the reference substituent. As it is capable of acting as both a lone pair hydrogen acceptor and a hydrogen donor in hydrogen bonding it is well suited for parameterization of media effects. The correlation equation used in this method is obtained by simply adding the term $Z\zeta$ to the correlation equation used for the subsets. Based on our results above the general correlation equation may therefore be written as:

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + A\alpha_X + M\mu_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + S\psi_X + Z\zeta_M + B^o \qquad (27)$$

The data for all of the sets studied is given in Table 3. The segmental method was again used as the steric parameterization. Values of the transport parameter τ_W for 4 - substituted benzenesulfonamides (set 31) were correlated with equation 27. τ_W is defined in a manner analogous to the definition of π by the equation:

$$\tau_{WX} = \log k_{W,XZ} - \log k_{W,HZ} \qquad (28)$$

where k_W is obtained by extrapolating to pure water the capacity factor k' determined in water -methanol or water - acetonitrile mixtures by reversed phase high performance liquid chromatography. X is the substituent of interest, and Z is either a constant substituent or H.
The best regression equation obtained is:

$$\tau_{WX} = 0.779(\pm 0.284)\sigma_{lX} + 4.35(\pm 0.997)\sigma_{eX} + 8.62(\pm 0.490)\alpha_X - 0.148(\pm 0.0384)\mu_X - 0.360(\pm 0.0559)n_{HX} + 0.191(\pm 0.0830) \qquad (29)$$

100R^2, 95.52; Adj. 100R^2, 94.98; F, 136.4; S_{est}, 0.189; S°, 0.231; n, 38.
C_l, 22.6; C_e, 12.6; C_α, 50.0; C_μ, 4.30; C_{nH}, 10.5.

In this set the term in ζ is not significant. As the k_W values used to define τ have been extrapolated to pure water this is not surprising.

Values of log k' for 3 - and 4 - substituted phenylacetylanilides (sets 32m and 32p respectively) in 50 % v/v aqueous methanol, 70 % v/v aqueous methanol, and (40 - 5 - 55) % v/v methanol - tetrahydrofuran - water were correlated with equation 27. The best regression equations are for set 32m:

$$Logk'_X = 1.60(\pm 0.355)\sigma_{lX} + 0.402(\pm 0.137)\sigma_{dX} - 2.60(\pm 0.800)\sigma_{eX} - 0.318(\pm 0.0702)\mu_X - 0.309(\pm 0.0659)n_{HX} - 0.0760(\pm 0.0197)n_{nX} + 0.610(\pm 0.115)\upsilon_{1X} + 0.310(\pm 0.123)\upsilon_{2X} + 1.14(\pm 0.0459)\zeta_M + 0.363(\pm 0.0653) \qquad (30)$$

100R^2, 95.17; Adj. 100R^2, 94.37; F, 103.0; S_{est}, 0.134; S°, 0.242; n, 57.
C_l, 37.8; C_d, 9.51; C_e, 6.18; C_μ, 7.53; $C_{\upsilon 1}$, 14.4; $C_{\upsilon 2}$, 7.34; C_{nH}, 7.31; C_{nn}, 1.80; C_ζ, 8.08.
and for set 32p:

$$Logk'_X = 16.0(\pm 3.74)\sigma_{lX} + 17.5(\pm 4.63)\sigma_{dX} - 90.5(\pm 27.8)\sigma_{eX} - 44.5(\pm 10.2)\alpha_X - 5.35(\pm 1.34)\mu_X + 7.72(\pm 2.41)n_{HX} + 1.23(\pm 0.341)n_{nX} + 18.4(\pm 4.75)\upsilon_{1X} + 24.0(\pm 6.73)\upsilon_{2X} + 1.18(\pm 0.0332)\zeta_M + 0.518(\pm 0.0386) \qquad (31)$$

100R^2, 98.97; Adj. 100R^2, 98.51; F, 182.5; S_{est}, 0.0665; S°, 0.128; n, 30.
C_l, 14.7; C_d, 16.1; C_e, 8.34; C_α, 8.20; C_μ, 4.93; $C_{\upsilon 1}$, 17.0; $C_{\upsilon 2}$, 22.1; C_{nH}, 7.11; C_{nn}, 1.14; C_ζ, 0.325.

We suspect that equation 31 is an artifact as its coefficients other than Z are so very much larger than those of equation 30.

Values of log k' for substituted benzenes in these same solvent mixtures (set 33) were also correlated with equation 27 giving the regression equation:

$$\log k'_{PhX} = 1.50(\pm 0.461)\sigma_{lX} + 0.402(\pm 0.187)\sigma_{dX} - 4.49(\pm 1.14)\sigma_{eX} - 0.428(\pm 0.0936)\mu_X - 0.489(\pm 0.079)n_{HX} - 0.0948(\pm 0.026)n_{nX} + 0.626(\pm 0.158)\upsilon_{1X} + 1.15(\pm 0.091)\zeta_M + 0.776(\pm 0.090) \qquad (32)$$

100R^2, 90.61; Adj. 100R^2, 89.08; F, 50.67; S_{est}, 0.180; S°, 0.338; n, 51.
C_l, 34.7; C_d, 9.30; C_e, 10.3; C_μ, 9.86; $C_{\upsilon 1}$, 14.4; C_{nH}, 11.3; C_{nn}, 2.18; C_ζ, 7.95.

Correlation with equation 26 of log k' values for substituted benzenes in aqueous methanol, acetonitrile, and tetrahydrofuran as the mobile phase (set 34) gave as the best regression equation:

$$\log k'_{PhX} = 1.21(\pm 0.391)\sigma_{lX} + 4.32(\pm 1.72)\sigma_{eX} + 6.40(\pm 0.801)\alpha_X - 0.389(\pm 0.0660)\mu_X - 0.837(\pm 0.130)n_{HX} + 1.06(\pm 0.242)\upsilon_{1X} + 0.354(\pm 0.138)\zeta_M + 1.19(\pm 0.162) \qquad (33)$$

$100R^2$, 87.73; Adj. $100R^2$, 86.35; F, 53.14; S_{est}, 0.280; S°, 0.376; n, 60.
C_l, 22.7; C_e, 8.14; C_α, 24.1; C_μ, 7.33; $C_{\upsilon 1}$, 19.9; C_{nH}, 15.8; C_ζ, 2.00.

Finally, we have correlated with equation 27 log k' values for a set of substituted benzenes in which the medium variation is in the column packing (set 35). The bonded phases studied were pentafluorophenyl dimethyl silane, heptafluoroisopropoxypropyl dimethyl silane, heptadecafluorodecyl dimethyl silane, and decyl dimethyl silane. The value for X = Bu when the bonded phase is decyl dimethyl silane was excluded as an outlier. The best regression equation obtained was:

$$k'_{PhX} = 0.303(\pm 0.143)\sigma_{lX} + 1.52(\pm 0.322)\alpha_X - 0.165(\pm 0.0225)\mu_X - 0.341(\pm 0.0404)n_{HX} + 0.279(\pm 0.0998)\upsilon_{1X} + 1.27(\pm 0.0743)\zeta_M - 0.805(\pm 0.108)i_X + 0.725(\pm 0.0790) \quad (34)$$

$100R^2$, 88.87; Adj. $100R^2$, 88.07; F, 93.56; S_{est}, 0.175; S°, 0.349; n, 90.
C_l, 11.8; C_α, 11.8; C_μ, 6.38; $C_{\upsilon 1}$, 10.8; C_{nH}, 13.2; C_i, 31.2; C_ζ, 14.8.

These results demonstrate the validity of the application of the Zeta method to transport parameters. They also provide further evidence of the importance of μ as a parameter and the utility of the segmental parameterization of the steric effect.

Discussion

Parameters. The combination of the use of the dipole moment as a parameter and the use of the segmental model for parameterizing the steric effect have resulted in a very dramatic improvement in the quality of the model. The dipole moment should be required as a parameter whenever the permanent dipole moment of the unsubstituted substrate is significantly less than that of the largest moment of opposite sign of any substituent in the data set. This is a consequence of the additivity of group moments. It is not surprising that the dipole moment gives somewhat better results than does its square. As was noted above the use of the latter was derived for the case of random orientations of the interacting molecules. If the interaction is between oriented molecules μ is the preferred parameter.

The dipole moments used for all of the members of set 12 were those of the corresponding PhX. The justification for this was the assumption that the actual variable was the C[sp^2]-X bond moment of which the PhX dipole moment is a measure. An analogous approach was applied to sets 12 and 21. The required parameter was assumed to be the bond moment of the C[sp^3]-X bond of which the MeX dipole moments are a measure, and were therefore used as parameters. The success of the model supports the assumption that the bond moment is the required parameter is justified.

The need for a steric effect parameterization presumably results from different steric effects on solvation of the substituted substrate in the initial and final phases. It is noteworthy that generally in sets where the substituent is bonded to an sp^2 hybridized carbon atom the steric effect occurs only at the first segment. By contrast, in sets where the substituent is bonded to an sp^3 hybridized carbon atom no steric effect occurs at the first segment.

The n_H and i parameters seem to be much more effective than the a_H and i° parameters respectively. We believe that in the case of the a_H parameter this is due to the existence of a leveling effect in the systems studied. When either the initial or the final phase is aqueous the concentration of water is not less than 55 molar. In this phase any substrate-water hydrogen bond that can form will do so. Parameters which are derived from equilibrium constants measured for dilute solutions in nonpolar solvents will not be useful in modelling the hydrogen bonding in an aqueous phase.

There is a hydrogen bonding problem which we have not yet addressed entirely. The aromatic ring is capable of hydrogen bonding.

Some or all of the electrical effect parameters are significant in all of the data sets studied. Their purpose is most likely to account for fact that there is a solvent effect on the dipole moment. The group dipole moments we have used as parameters were obtained from gas phase data when possible, otherwise they were obtained in nonpolar solvents. The values of the group moments in water are probably very different from these. When the G group of XG is composed of sp^2 hybridized C atoms the electrical effect parameters may also account for the effect of X on charge transfer and hydrogen bond acceptor interactions of G.

We may obtain further insight into the parameterization by comparing predicted and observed signs of the coefficients of the regression equations obtained for correlations of log P values. In partition studies water is the initial phase and the organic solvent is the final phase. The interaction between the substrate and the initial phase should be given by:

$$Q_X = L_W\sigma_{lX} + D_W\sigma_{dX} + R_W\sigma_{eX} + M_W\mu_X + A_W\alpha_X + H_{1W}n_{HX} + H_{2W}n_{nX} + I_Wi_X + B_{DW}n_{DX} + B_{AW}n_{AX} + S_W\psi_X + B_W^o \qquad (35)$$

while that in the final phase should be given by:

$$Q_X = L_O\sigma_{lX} + D_O\sigma_{dX} + R_O\sigma_{eX} + M_O\mu_X + A_O\alpha_X + H_{1O}n_{HX} + H_{2O}n_{nX} + I_Oi_X + B_{DO}n_{DX} + B_{AO}n_{AX} + S_O\psi_X + B_O^o \qquad (36)$$

where W designates the aqueous phase and O the organic phase. Then from equation 1:

$$Q_X = (L_O-L_W)\sigma_{lX}+(D_O-D_W)\sigma_{dX}+(R_O - R_W)\sigma_{eX}+(M_O-M_W)\mu_X+(A_O-A_W)\alpha_X +(H_{1O}-H_{1W})n_{HX}+(H_{2O}-H_{2W})n_{nX}+(I_O-I_W)i_X+(B_{DO}-B_{DW})n_{DX} +(B_{AO}-B_{AW})n_{AX}+(S_O-S_W)\psi_X+(B_O^o-B_W^o) \qquad (37)$$

Thus, any coefficient V of an independent variable v in the IMF equation is given by:

$$V = V_O - V_W \qquad (38)$$

Properties of a number of solvents are given in Table 4. They include values of α, μ, n_H, n_n, and the molarity of the pure solvent, M_{Sv}. Also tabulated are values of $M_{Sv}\alpha$, $M_{Sv}\mu$, $M_{Sv}n_H$, and $M_{Sv}n_n$. These values suggest that H_{1W}, H_{2W}, and M_W are much larger than H_{1O}, H_{2O}, and M_O. It follows then that the signs of the coefficients H_1, H_2, and M should be negative in all of the sets studied. This is indeed the case. Polarizability should be much more important in the organic phase than in the aqueous phase. This should also be true of the coefficient B_{CG} which is also a measure of polarizability. Thus the values of A and B_{CY} should all be positive. Again this is the case in all the sets studied. It is interesting to note that steric effect coefficients S_i are all positive for these sets. This presumably results from steric shielding of the substituent X by the group Y which decreases the extent of hydrogen bonding and of dipole - dipole and dipole -induced dipole interactions in the water phase, resulting in an increased concentration of substrate in the organic phase. L, D, and R, the coefficients of the electrical effect parameters are also positive in all cases. Why this should be the case is unclear.

Table 4. Solvent properties.

Solvent	M_{Sv}	n_H	n_n	μ	α	M_W *in Sv*[a]
water	55.6	2	2	1.88	0.018	---
c-hexane	9.25	0	0	0	0.257	0.00245
chloroform	12.4	1	0	1.04	0.191	0.067
PGDP	2.56	0	4	2.4[b]	1.012	0.665
1-octanol	6.35	1	1	1.62	0.390	1.72
hexane	7.66	0	0	0	0.278	0.00406

Solvent	M_{Sv} *in W*[a]	$M_{Sv}n_H$	$M_{Sv}n_n$	$M_{Sv}\mu$	$M_{Sv}\alpha$
water	---	111	111	105	1.00
c-hexane	0.00065	0	0	0	2.38
chloroform	0.068	12.4	0	12.9	2.37
PGDP	---	0	10.2	6.1	2.59
1-octanol	0.00404	6.35	12.7	10.3	2.48
hexane	0.00011	0	0	0	2.13

Abbreviations: C, molarity;W, water; Sv, solvent; c, cyclo; PGDP, propylene glycol dipelargonate.

a. From ref. 7. b. Estimated form the dipole moments of ethylene glycol diacetate and meso - 2,3 - diacetoxybutane.

We may apply the same type of analysis to the regression equations obtained for the correlation of log k' values (sets 31, 32m, 33 - 35). The final phase in these sets is the column surface, the initial phase is an aqueous or aqueous - organic phase. Again we predict negative signs for the coefficients of n_H, n_n, and μ; and a positive sign for the coefficients of α and υ_1. These predictions are borne out by the results. Again, the coefficients of the electrical effect parameters are also positive in all cases. The agreement between predicted and observed signs for the coefficients of regression equations obtained by correlation of log P and log k' values with the IMF equation lends further support to the conclusion that transport parameters are composite quantities which are a function of the difference in intermolecular forces between an initial and a final phase.

Parameter Contributions. We now consider the relative contributions of the various parameters. Values of C_i for sets where G is either Ph or is an sp^2 hybridezd group are given in Table 5. The C_i values for set 12 were recalculated excluding C_{nCG}, those for sets 33, 34, and 35 were recalculated excluding C_ζ

Table 5. Values of C_i for PhX and for GX hybridized sp^2.(set 12).

Set	*1*	*2*	*12*	*31*	*33*	*34*	*35*
σ_l	34.9	29.0	33.7	22.6	37.7	23.2	13.8
σ_d	0	0	0	0	10.1	0	0
σ_e	4.26	0	4.13	12.6	11.2	8.31	0
μ	8.06	13.2	8.66	4.30	10.7	7.48	7.49
α	23.8	20.6	27.3	50.0	0	24.6	13.8
n_H	6.56	12.1	8.17	10.5	12.3	16.1	15.5
n_n	3.06	3.02	3.83	0	2.37	0	0
i	0	8.16	0	0	0	0	36.6
υ_1	12.9	13.9	14.2	0	15.6	20.3	12.7
υ_2	*6.47*	0	0	0	0	0	0

Values in italics are for parameters that were significant at the 90.0 % confidence level.

It is clear that with the exception of the borderline dependence on υ_2 the C_i values for set 12 are essentially the same as those for set 1. They are different however from those of the remaining sets. All of the sets studied are dependent on σ_l, μ, and n_H. Allowing for collinearities between α and υ_1 it is highly likely that they are all dependent on these variables as well. There may also be a dependence on σ_e and n_n .

Values of C_i for sets in which G is alkyl or in which the C atom that X is bonded to is hybridized sp^3 are given in Table 6. The C_i values for sets 11 and 21 have been recalculated after the exclusion of the contribution for n_{CG}

Hydrogen Bonding for the Ethynyl, Substituted Ethynyl, and Thiol Groups. Calculation of Log P values for the ethynyl, phenylethynyl, and thiol groups from equation 13 and comparison with the observed values shows that the ethynyl group can act as both hydrogen donor and acceptor. The phenylethynyl group can act as a hydrogen acceptor. It seems likely though it is as yet unproven that the carbon carbon triple bond can generally act as a hydrogen acceptor. The results for the thiol group are inconclusive. It should be noted that the use of $n_H = 1$ for the groups C_2H and CH_2C_2H and $n_n = 1$ for these groups and C_2CPh groups in sets 3, 11, and 12 gave good results supporting the conclusion that these groups are involved in hydrogen bonding.

Table 6. Values of C_i for PhX and AkX.

Set	*1*	*2*	*3*	*4*	*11*	*21*
σ_l	34.9	29	0	0	21.1	22.7
σ_d	0	0	12.1	16.1	11.5	14.3
σ_e	4.26	0	12.7	13.5	8.55	8.57
μ	8.06	13.2	12.2	10.3	10.5	10.3
α	23.8	20.6	44.9	43.0	28.2	32.9
n_H	6.56	12.1	10.7	0	7.06	6.12
n_n	3.06	3.02	7.48	6.34	4.93	5.02
i	0	8.16	0	10.7	0	0
υ_l	12.9	13.9	0	0	0	0
υ_2	*6.47*	0	0	0	8.17	0

All of the alkyl sets depend on σ_d, σ_e, μ, α, and n_n. In view of the collinearity between n_H and i they probably depend on the former as well.

The Nature of Transport Parameters. Our present results provide further support for our previous conclusion that transport parameters are composite and represent the difference in intermolecular forces between an initial and a final phase. The relationship between these parameters and hydrophobicity (or lipophilicity) depends only on the difference between intermolecular forces in the initial and the final phases. **There is no special hydrophobic effect.** The observed effects are a natural result of the difference in intermolecular forces in initial and final media.

The Hansch - Fujita Model. It has been established that as long as all the necessary pure parameters are included in the composite parameters used, a model constructed from composite parameters is completely equivalent to one which uses pure parameters in representing the data.[13] The only reason for using pure parameters lies in the ease of interpretation of the results. In its use of transport parameters, in particular log P or π, the Hansch- Fujita (HF) model uses composite parameters.

As we have pointed out earlier the Hansch model often requires in addition to transport parameters the use of electrical effect[14-17], steric effect[14-17], and polarizability[14-17] parameters. It sometimes requires dipole moment[18] parameters as well. The need for these parameters results from the fact that the mix of pure parameters found in a particular transport parameter may not be that which is required for a particular type of bioactivity. This is hardly surprising. **The probability that all biomembranes and all receptor sites will require the same mix of pure parameters is extremely small.** The addition of electrical, steric and polarizability terms provides the proper parameter composition for modelling the bioactivity being studied. To illustrate the point let us consider a typical HF correlation equation:

$$ba_X = T\tau_X + \rho\sigma_X + AMR_X + S\upsilon + B^o \qquad (39)$$

where ba is the bioactivity; σ is a composite electrical effect parameter of the Hammett type; τ is a transport parameter such as log P, π, or log k'; MR is the group molar refractivity; υ is a steric parameter; and T, ρ, A, S, and B°, are coefficients.
We have shown that σ is given by the expression:

$$\sigma_X = l\sigma_{lX} + d\sigma_{dX} + r\sigma_{eX} + h \qquad (40)$$

From equation 3 we obtain:

$$MR_X = 100(\alpha_X + 0.0103) = 100\alpha_X + 1.03 \qquad (41)$$

From equation 5 we have:

$$S\upsilon = S_1^*\upsilon_1 + S_2^*\upsilon_2 + S_3^*\upsilon_3 + S_o^* \qquad (42)$$

finally, we have shown that τ is given by the equation:

$$\tau_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + M\mu_X + S_1\upsilon_{1X} + S_2\upsilon_{2X} + S_3\upsilon_{3X} + B^o \qquad (43)$$

On substituting equations 40 through 43 into equation 39 we have:

$$\log ba_X = (L + \rho l)\sigma_{lX} + (D + \rho d)\sigma_{dX} + (R + \rho r)\sigma_{eX} + (A + 100A^*)\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + M\mu_X + (S_1 + SS_1^*)\upsilon_{1X} + (S_2 + SS_2^*)\upsilon_{2X} + (S_3 + SS_3^*)\upsilon_{3X} + B^o + \rho h_o + 1.03A^* + SS_o^* \qquad (44)$$

Equation 44 may be rewritten as:

$$\log ba_X = L'\sigma_{lX} + D'\sigma_{dX} + R'\sigma_{eX} + A'\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + M\mu_X + S_1'\upsilon_{1X} + S_2'\upsilon_{2X} + S_3'\upsilon_{3X} + B'o \qquad (45)$$

which is a form of the IMF equation. It follows then that the bioactivity is a function of the difference in intermolecular forces between initial and final states. This does not mean that transport parameters should not continue to be used in modelling bioactivities. It simply provides an explanation of why and how they work. There is a very important point which must be made here. As any combination of pure and/or composite parameters which has the correct composition will serve to quantitatively describe a phenomenon **it is not necessary to use log P or π values.** Bioactivities can be correlated directly with the IMF equation.

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Contributions from Individual Intermolecular Forces. The intermolecular force model cannot be used to identify the contributions of individual imf because some imf are a function of the same parameters. Thus dd, di and Id interactions are all a function of the bond moment, while di, ii, and Ii interactions are all a function of polarizability. A second problem is that there are significant collinearities between some of the parameters. Thus, the bond moment μ, the localized electrical effect parameter σ_l and to a lesser extent the hydrogen bonding parameter n_n are all collinear. This has no effect on the validity of the imf model.

Conclusions

The intermolecular force model provides a much improved quantitative description of transport parameters for substituted benzenes when the bond moment μ is included as a parameter and the steric effect is modeled by the multiparametric segmental method. The model may undergo slight further improvement when υ_1 is replaced by υ_Δ. Use of μ^2 in place of μ gave poorer results. We have also shown why μ cannot in most cases be represented by electrical effect parameters. Attempts at improving the parameterization for hydrogen bonding and for charged substituents were unsuccessful. In parameterizing the hydrogen bonding it is necessary to take into account the levelling effect in aqueous phases which results in the formation of the greatest possible number of hydrogen bonds between water and substrate. The general form of the IMF equation may now be written as:

$$\begin{aligned} Q_X = {} & L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + M\mu_X + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + I i_X \\ & + B_{DX} n_{DX} + B_{AX} n_{AX} + S\psi_X + B^{o} \end{aligned} \qquad (46)$$

The intermolecular force model provides an explanation for the success of the Hansch model in the quantitative description of bioactivities.

Acknowledgement

The author would like to acknowledge helpful discussion with P. J. Taylor, and the gift of computer hardware from Magical Designs Inc.

Literature Cited

1. Charton, M., *Topics in Current Chem.*, 114, 107-118 (1983).
2. Charton, M., in *QSAR in Design of Bioactive Compounds*, M. Kuchar (Ed.), J.R. Prous, Barcelona, 1985, pp. 41-51.
3. Charton, M., in *Rational Approaches to the Synthesis of Pesticides*, P.S. Magee, J.J. Menn and G.K. Koan (Eds.), American Chemical Society, Washington, D.C., 1984, pp. 247-278.
4. Charton, M., in *Trends in Medicinal Chemistry '88*, H. van der Goot, G. Domany, L. Pallos and H. Timmerman (Eds.), Elsevier, Amsterdam, 1989, pp. 89-108.
5. Charton, M., *Prog. Phys. Org. Chem.* **18**, 163-284 (1990).
6. Charton, M., *Stud. Org. Chem.*, **42**, 629-687 (1992).

7. Abraham, M. H., Duce, P. P., Porter, D. V., Barratt, D. G., Morris, J. J., and Taylor, P. J., *J. Chem. Soc. Perkin Trans. II*, 1355-1375 (1989).

8. Charton, M., *Prog. Phys. Org. Chem.*, **16**, 287-315 (1987).

9. Charton, M., in *The Chemistry of the Functional Groups. Supplement A, The Chemistry of Double Bonded Functional Groups.*, Vol. **2**, Part 1, S. Patal and Z. Rappaport (Eds.), Wiley, New York, 1989, pp. 239-298.

10. Charton, M., in *The Chemistry of Sulfenic Acids, Esters and Derivatives*, S. Patai, ed., Wiley, New York, 1990, pp.657-700.

11. McClellan, A. L., *Tables of Experimental Dipole Moments*, W. H. Freeman, San Francisco, 1963.

12. McClellan, A. L., *Tables of Experimental Dipole Moments Vol. 2*, Rahara Enterprises, El Cerrito, Cal., 1974.

13. Charton, M., Greenberg, A., and Stevenson, T. A., *J. Org. Chem.* **50**, 2643-2646 (1985).

14. Martin, Y. C., *Quantitative Drug Design*, Dekker, New York, 1978.

15. Franke, R., *Theoretical Drug Design Methods*, Akademie - Verlag, Berlin, 1984.

16. Gupta, S. P., *Chem. Rev.* **87**, 1183 - 1253 (1987).

17. Hansch, C., *Drug Developement Res.* **1**, 267 - 309 (1981).

18. Lien, E. J., Guo, Z. - R., Li, R. - L., and Su, C. - T., *J. Pharm. Sci.*, **71**, 641-655 (1982).

RECEIVED April 26, 1995

Chapter 7

QSARs from Mathematical Models

Systemic Behavior of Pesticides

D. A. Kleier

Stine-Haskell Research Center, DuPont Agricultural Products,
P.O. Box 30, Newark, DE 19714

Quantitative structure-activity relationships (QSAR's) are usually established by a statistical analysis of a matrix of property data for a series of compounds. QSAR's can also be established by experimentally validating a mathematical model. Such a first principles model has been developed for phloem systemicity of a compound as a function of its physical properties such as partition coefficients and acid dissociation constants. This model has been validated by both systematic experimental studies, and general observations concerning the phloem systemicity of pesticides. The mathematical model can account for the sensitivity of systemicity to plant parameters and has recently been used to design a phloem systemic pro-nematicide.

Mathematical modeling of an idealized system can provide simple relationships that capture the essence of the more complex, real world system being modeled. We have developed such a relationship for the movement of both ionizable and non-ionizable compounds in the plant vascular system (*1*). This relationship enables a deeper understanding of the delicate balance of physical properties required for phloem systemicity. Phloem systemic compounds move in that part of the plant vascular system known as the sieve tubes (*2*). Phloem sap moving through these tubes carries photosynthates, especially sucrose, from mature leaves to sinks in both the roots and in meristematic tissue. Compounds that can enter the phloem and be retained there are transported along with these photosynthates.

What is the value of a phloem systemic pesticide? First of all, phloem systemic insecticides and fungicides can control pests far from the site of a foliar application (*3*). Furthermore, since phloem systemic pesticides would normally be applied foliarly, their performance should not depend upon soil conditions, nor should they require rain for activation. They should also have less chance of leaching into ground water. Such compounds may also provide more flexibility to the farmer. For example, the need for

0097–6156/95/0606–0098$12.00/0

prophylactic soil treatments should be lessened by the availability of a phloem systemic curative material that could be applied to the leaf canopy as needed.

Over the past several years a great deal of information has been collected concerning the relationship of chemical structure to phloem mobility *(4-10)*. First of all, the vast majority of compounds known to be phloem mobile are acidic in nature. These materials include benzoic acids, naphthoxyacetic acids, sulfonylureas and maleic hydrazide to name just a few. The phloem mobility of weak acids has been attributed *(11)* to entrapment of the dissociated acids in the phloem sap which is relatively basic compared to the surrounding tissue. Any QSAR model should take this effect into account. There are a few compounds known to be phloem mobile that are not acidic. These include herbicides such as amitrole and some quaternized nitrogen heterocycles such as N-methylpyridinium halides (8,*10*). Most, if not all, non-acidic, phloem mobile compounds have $\log K_{OW}$ values of zero or below. We set out to develop a quantitative structure-activity relationship that would not only explain the phloem mobility of both acidic and hydrophilic compounds as a function of physical properties, but would also be sensitive to the type and growth stage of the plant to which the compounds had been applied (*1*).

QSAR's are traditionally developed by establishing an empirical relationship between some descriptor of biological activity and multiple descriptors of chemical structure. Empirical QSAR's start with experimental data and use statistical methods such as linear regression to establish the relationship. The resultant relationship should be consistent with fundamental laws. Alternatively, QSAR's can be constructed starting from these fundamental laws. This deterministic or ab initio approach starts with a physical model of a biological system, then uses basic physico-chemical laws to develop a structure-activity relationship. QSAR's developed in this manner should be validated by experiments. Our QSAR for phloem mobility was developed using a mixed or semi-empirical method. We first developed a deterministic model for the plant vascular system, and then used a statistical method to estimate some of the key parameters needed to apply the model in a practical sort of way.

In this paper we will first briefly review our simple model for the plant vascular system. We will then describe the QSAR that results from a mathematical development of this model. The relationship relates mobility to a compound's acid dissociation constant as well as to its octanol-water partition coefficients. We will then review the performance of the model for pesticides. Finally, we will describe the application of the QSAR to the design of novel phloem mobile pesticides.

Description of the Model

A cross sectional view of the plant vascular system (*12*) reveals that the sieve tubes which transport phloem sap in one direction, and the xylem elements that carry the transpiration stream in the opposite direction are in intimate contact in most higher plants. As a result compounds moving in the phloem can be lost to the xylem and carried off in the opposite direction. This limits net transport for compounds that are highly permeable.

Our simple model of the plant vascular system captures the essence of these transport and permeation processes. Following Tyree, et al. (*7*) the model consists of two coaxial pipes as shown in Figure 1. The inner pipe of radius r represents the sieve tubes which carry phloem sap from the top of the model plant to the bottom. The outside pipe representing the apoplast carries xylem sap in the opposite direction. The pipes are separated by a permeable membrane. It is assumed that compound is applied to an annulus located at the top of the model plant. This annulus represents the apoplast of the leaf. The model provides for permeation of the compound from this annulus into the sieve tube which it surrounds. Once entrained, the model then provides for movement by mass flow of the phloem sap towards the bottom of the model plant. The model also provides for loss in transit via permeation back into the surrounding apoplast where the xenobiotic can be swept back to the leaf with the transpiration stream. The very important acid trapping effect is accounted for by a) assigning a higher pH to the phloem sap than to the surrounding apoplast, b) providing for correspondingly higher degrees of acid dissociation within the phloem consistent with the pH differential and c) assigning reduced permeabilities to charged conjugate bases resulting from acid dissociation (*1*). For most of the simulations performed we have taken the pH of the phloem to be 8 and that of the surrounding apoplast to be 6.

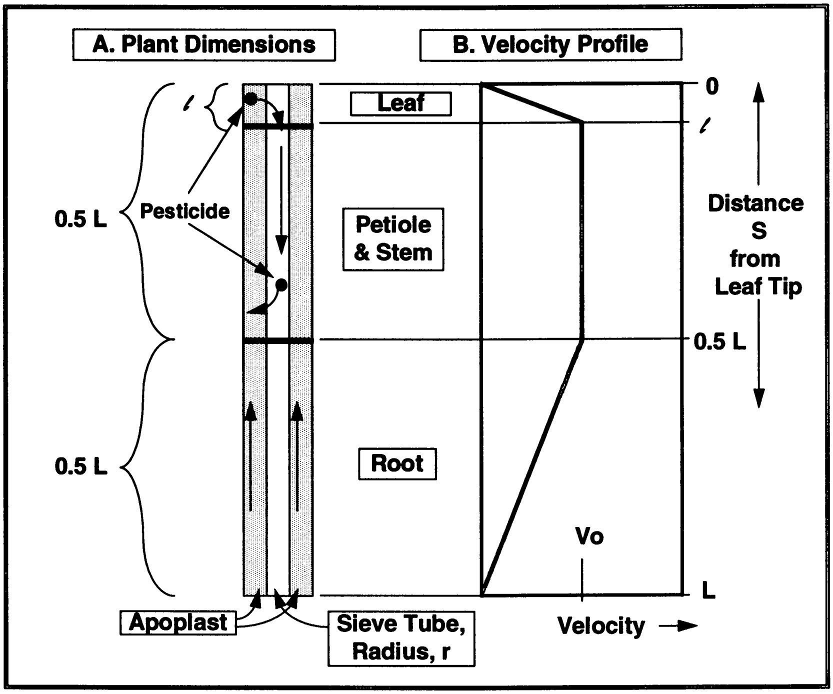


Figure 1. Simple model of plant vascular system (Reprinted with permission from Reference 13.)

The total length L of the plant is divided into three parts: leaf, petiole & stem, and root. The root length is taken as half that of the plant (L / 2). The upper half of the plant is divided into a leaf of length *l*, and a petiole and stem of length (L / 2) - *l*. What distinguishes these three parts of our model plant is the phloem sap velocity profile within them. The velocity at the leaf tip is taken as zero, and is assumed to rise linearly to a velocity v_o at the base of the leaf. Within the petiole and stem the phloem sap velocity is assumed to remain at a constant value of v_o. In the root we assume a linear decrease in velocity back to a value of zero at the root tip. We use the concentration factor, C_f, as the dependent variable of this model. It is defined as the stoichiometric concentration of xenobiotic in the phloem at a distance S from the leaf tip at time t relative to that in the leaf apoplast.

$$C_f(S,t) = \{[HA]_S + [A^-]_S\} / \{[HA]_o + [A^-]_o\} \qquad \text{(Equation 1)}$$

Here $[HA]_S$ and $[A^-]_S$ are respectively the concentrations of the undissociated acid HA and conjugate base A^- within the phloem sap at a distance s from the leaf tip. $[HA]_o$ and $[A^-]_o$ are the corresponding concentrations at the point of application within the leaf apoplast. Since the leaf apoplast is considered to be a reservoir of xenobiotic, $[HA]_o$ and $[A^-]_o$ are assumed independent of time. Large values of C_f at large distances s from the leaf tip indicate high phloem mobility.

The differential equations describing the movement of xenobiotic for this model are essentially statements of the conservation of mass of the xenobiotic. These equations account for both the downward mass flow along the axis of the sieve tube, as well as radial permeation of xenobiotic through the cylindrical surface of the sieve tube. An analytical solution to the differential equations describing the movement of xenobiotic is possible for this model and has been described elsewhere (*1,13*). The solution expresses the dependence of the concentration factor C_f upon both distance s and time t, and parametrically upon properties of both the plant and xenobiotic. According to this model a steady state is established throughout the length of the plant within a short time, usually within a few multiples of L / v_o (*13*). We have thus chosen to use the steady state concentration factor evaluated at a point deep within the root (S = 0.9L) as a measure of a compound's phloem mobility. In the steady state limit (t = ∞) the concentration factor evaluated at this point is given by

$$\begin{aligned} C_f(0.9L, \infty) = {} & \{([H^+]_i + K_a) / ([H^+]_o + K_a)\} \\ & \times \{5([H^+]_o p_{HA} + K_a p_{A^-})\} / \{[H^+]_i(b + p_{HA}) + K_a(b + p_{A^-})\} \\ & \times \exp\{-c([H^+]_i p_{HA} + K_a p_{A^-}) / ([H^+]_i + K_a)\} \end{aligned} \qquad \text{(Equation 2)}$$

where $b = rv_o / 2l$, and $c = (2.609L - 2l) / rv_o$. The plant parameters appearing in Equation 2 include the hydrogen ion concentrations within both the phloem sap and the apoplast, $[H^+]_i$ and $[H^+]_o$, respectively, as well as the length and velocity parameters illustrated in Figure 1. The xenobiotic parameters include the acid dissociation constant, K_a, as well as the permeability p_{HA} of the xenobiotic and that of its conjugate base p_{A^-}. The permeabilities are essentially rate constants for diffusion across the sieve tube membranes, and are very seldom reported in the literature. In order to make practical applications, we assume a log linear relationship between the permeability of a species and its octanol-water partition coefficient.

$$\log p_i = a \log K_{ow,i} + b \qquad \text{(Equation 3)}$$

The decreased permeability of a negatively charged conjugate base, A^-, relative to its undissociated parent acid, HA, is taken into account by using a log K_{ow,A^-} value in Equation 3 which is 3.7 log units less than log $K_{ow,HA}$ (*15*). The slope and intercept of this relationship are also plant parameters and could be related to membrane thickness and viscocity in a completely deterministic model (*14*). We have taken a more empirical approach at this point. The parameters in equation 3 have either been determined by linear regression of permeability data or simply assigned to account for the systemic properties of a set of xenobiotics. For example, the parameters for the so-called Nitella membranes ($a = 1.20$, $b = -5.86$) have been determined by linear regression using permeabilities measured (*16*) for a series of organic molecules through the cell membrane of the large single celled organism known by this name. The so-called Grayson parameters ($a = 1.20$, $b = -7.5$) have been assigned in order to provide a good account of the phloem mobility of a series of xenobiotics including a large number of pesticides (14).

Predictions of the Model: Dependence of the Concentration Factor, C_f, upon K_a and K_{ow}

In Figure 2 we display the dependence of the concentration factor evaluated using Equation 2 upon $K_{ow,HA}$ and K_a for a standard 15 cm. plant with Nitella membranes. Large values of log C_f correspond to high predicted mobility. Lipophilic, non-acidic compounds appear in the upper right hand corner of this contour surface. LogC_f values in this region are generally much less than -4. Thus, a herbicide like diuron with a LogK_{ow} = 2.8 and a pK_a > 12 is not predicted to move well in the phloem. As we move along the top edge of the plot from diuron towards less lipophilic, but still non-acidic compounds, phloem mobility is predicted to increase until an optimum is reached at logK_{ow} ~ -2.0. This is where marginally phloem mobile, non-dissociable compounds such as oxamyl and N-alkyl pyridiniums reside. The highest mobility is predicted for compounds that reside on the elevated diagonally running ridge near the center of the plot. On this ridge reside weak acids with pK_a values near 5 and logK_{ow} values near zero. Compounds that reside here have logC_f > 0, and hence are predicted to actually concentrate in the phloem at long distances from the point of application. This is the region where compounds known for their phloem mobility such as the sulfonylurea, metsulfuron-methyl, reside. It should be noted, however, that the model does not predict phloem mobility for all acids. Very lipophilic acids such as acifluorfen are not predicted by this model to be phloem mobile. While the general shape of this plot does not change with plant parameters, the position, breadth and height of the ridge are quite sensitive to parameters such as pH differential, membrane parameters, and plant dimensions.

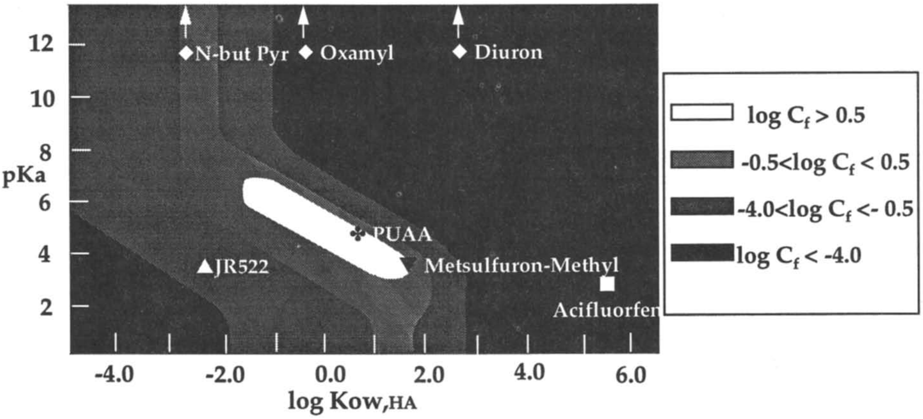


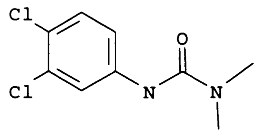
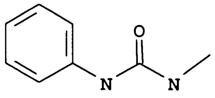
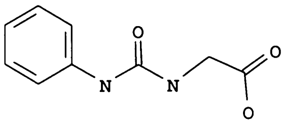
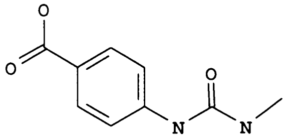
Figure 2. Contour plot for log Cf as a function of $\log K_{ow,HA}$ and pK_a. Plant parameters are L = 15 cm, l = 5 cm, r = 5 mm, v_o = 1.8 cm/min. The membrane parameters (Equation 3) are a = 1.20 and b = -5.86. The pH of the phloem sap was taken as 8.0 and that of the surrounding apoplast as 6.0. PUAA = phenylureidoacetic acid, N-but-Pyr = N-butylpyridinium bromide, JR522 is illustrated in Figure 3 (vide infra).

Tasteful Rules for Rendering Xenobiotics Phloem Systemic

How can the properties of a lipophilic compound such as diuron be altered in order to make it phloem mobile? Either the compound's $\log K_{ow}$ or pK_a or both can be lowered in order to move it into the loftier regions of the concentration factor contour diagram (Figure 2). Molecular modifications of diuron such as sugar conjugation (colloquially termed "sweetening") should lower the $\log K_{ow}$ and according to the model improve phloem mobility. Quaternization, or "salting", is another modification that significantly lowers the $\log K_{ow}$ of compounds possessing basic nitrogens, and, as suggested by studies on pyridiniums (*8, 10*), should improve phloem mobility. Diuron, which does not possess a nitrogen basic enough to be quaternized, cannot be "salted", but it can be "soured" by acid functionalization. For example, the addition of a carboxylic group is predicted to be a very effective modification for increasing the phloem mobility of phenylurea herbicides as shown in the Table I. While neither diuron nor defenuron is predicted to move well in the phloem, both acid derivatives, PUAA and CPMU, are predicted to have concentration factors greater than 1.0, and hence both compounds are expected to be highly phloem mobile in a plant with parameters similar to the one simulated. C_f values greater than 1.0 can be attributed to the acid trapping effect. The predictions of the table are consistent with studies performed on both one-leaf bean (*phaseolus vulgaris*) and mustard plants (*sinapis alba*) plants by Neumann, et al. (*17*). The expected high phloem mobility of PUAA is also indicated by its location on the elevated ridge in Figure 2. Unfortunately, the acid derivatives of defenuron are devoid of herbicidal activity.

In fact, it is anticipated that loss of useful intrinsic biological activity will be a rather general problem for pesticides that have been modified to improve phloem systemicity. The physical properties required for pesticidal activity are often incompatible with those required for phloem systemicity. One possible solution to this

Table I. Physical Properties and Theoretical[a] Concentration Factors for Phenylureas

Structure	Name	$\log K_{ow}$	pK_a	C_f
	Diuron	2.8	>12	0.00
	Defenuron	1.1	>12	0.00
	Phenylureido Acetic Acid (PUAA)	0.77	4.76	36.1
	4-Carboxy-phenyl, methylurea (CPMU)	1.10	4.2	32.4

[a] Plant parameters for these simulations are identical to those described in the caption to Figure 2.

quandary is the use of phloem mobile pro-pesticides. For example, if a substituent (e.g., acid group, sugar) that facilitates movement in the phloem could be removed near the site of action, effective control might be realized. Realization of this stragegy might necessitate modifying not only the crop protection chemical but also the crop to be protected. The modification of the chemical would be designed to enhance phloem systemicity; that of the plant to ensure activation at the site where protection is needed.

The Pro-pesticide Approach to the Design of Phloem Systemic Nematicides

A highly phloem mobile nematicide would be a very useful crop protection chemical. Such a compound could be sprayed on the leaves of a plant and still control nematodes that feed on the roots. This was the motivation behind the design and synthesis (*18*) of the following derivative of oxamyl:

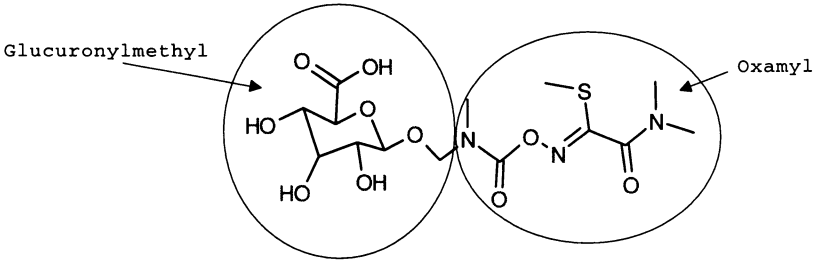


Figure 3. Structure of JR522.

The right hand side of this structure is oxamyl itself, a known effective nematistat. The substituent on the left is a glucuronylmethyl group which is predicted to improve the mobility of oxamyl by lowering both its pK_a and $\log K_{ow}$. This mobilophore consists of both a sugar and an acid functionality and hence might be termed "sweet and sour". Figure 2 illustrates the location of both oxamyl and the glucruonide (designated as JR522) on the concentration factor surface for a short plant with Nitella membranes. For this simulated plant oxamyl falls in a region that is not predicted to be particularly well suited for phloem translocation ($\log C_f < -4$). On the other hand the glucuronide with lowered pK_a and $\log K_{ow}$ is predicted to lie on a contour with a $\log C_f$ value of between -3 and -4. Our experience has been that compounds with predicted $\log C_f$ values in this region usually exhibit phloem systemicity. Furthermore, since the glucuronide resides on a less steep portion of the C_f surface than does oxamyl, its predicted mobility will be less sensitive to changes in plant parameters.

Experiments with transgenic tobacco plants that have been engineered to hydrolyze the glucuronide in a root specific manner (*18, 19*) are consistent with the expectation that JR522 will be phloem mobile. At a foliar application rate of 1 gm/l the glucruonide provides significantly greater control than oxamyl itself when applied at over 10 gm/l. We believe this to be due to the enhanced phloem mobility of the glucuronide coupled with root specific generation of the active substance near the target pest.

Summary

A simple model of the plant vascular system has been developed which enables the prediction of the phloem mobility of xenobiotics as a function of their physical properties in a variety of plant types. The predictions of this model are consistent with the vast majority of reports on this subject. The model can be used to design phloem systemic xenobiotics including phloem systemic pro-pesticides.

Acknowledgments

The author is grateful to Dr. Francis Hsu for introducing the author to this area of research, for his efforts to validate the mathematical model, and for his promotion of modeling as a tool for the design and discovery of new pesticides. The author is grateful to Dr. King Mo Sun for his artful synthesis of the glucuronide of hydroxymethyloxamyl. I would also like to thank Dr. Tariq Andrea for useful discussions on the mathematical development of the simple model of the plant vascular system. Finally, I would like to thank former colleagues, James R. Sanborn and B. Terence Grayson, as well as my managers, John B. Carr, Russ F. Bellina and James V. Hay for their encouragement to pursue this area of research.

References

1. Kleier, D. A. *Plant Physiol.* **1988,** *86,* 803-810.
2. Grimm, E. and Neumann, S. In *Modern Selective Fungicides-Properties, Applications and Mechanisms of Action;* Lyr, H., Ed.; Longman Scientific & Technical: Essex, England, 1987; pp 13-29.
3. Edgington, L. V., *Ann. Rev. Phytopathol.* **1981,** *19,* 107-24.
4. Bromilow, R.; Rigitano, R. L. O.; Briggs, G. G.; Chamberlain, K. *Pestic. Sci.* **1987**, *19,* 85-99.
5. Rigitano, R. L. O.; Bromilow, R. H.; Briggs, G. G.; Chamberlain, K. *Pestic. Sci.* **1987**, *19,* 113-133.
6. Chamberlain, K.; Butcher, D. N.; White, J. C. *Pestic. Sci.* **1986**, *17,* 48-52.
7. Tyree, M.; Peterson, C. A.; Edgington, L. *Plant Physiol.* **1979**, *63,* 367-374.
8. Hsu, F.; Kleier, D. A. *Plant Physiol.* **1986,** *86,* 811-816.
9. Lichtner, F. In *Phloem Transport;* Cronshaw, J.; Lucas, W. J.; Giaquinta, R. J. Eds.; Alan R. Liss, New York, 1986, pp 601-608.
10. Price, C. E.; Boatman, S. G.; Boddy, B. J. *J. Exp. Botany,***1975**, *26,* 521-532.
11. Crisp, C. *Pestic. Chem.,* **1972**, *1*, 211-264.
12. Kuhn, W. *Angewandte Chem, Int. Ed., Eng.* **1990,** *29*, 1-19.
13. Kleier, D. A. *Pestic. Sci.* **1994,** *42,* 1-11.
14. Grayson, B. T.; Kleier, D. A. *Pestic. Sci.* **1990**, *30,* 67-79.
15. Scherrer, R. A. In *Pesticide Synthesis through Rational Approaches,* Magee, P. S.; Kohn, G. K.; Menn, J. J. Eds.; ACS Symposium Series, American Chemical Society, Washington, D. C. 1984, *No. 255,* 225-246.
16. Collander, R., *Physiol. Planta* **1954**, *7,* 420-445.
17. Neumann, S.; Grimm, E.; Jacob, F. *Biochem. Physiol. Pflanzen,* **1985**, *180,* 257-268.
18. Hsu, F.; Sun, K. M.; Kleier, D. A.; Fielding, M., *Pestic. Sci.,* **1995**, in press.
19. Yamamoto, Y. T.; Taylor, C. G.; Acedo, G. N.; Cheng, C. L.; Conkling, M. A. *Plant Cell,* **1991**, 3, 371-382.

RECEIVED April 26, 1995

Chapter 8

Hydrophobicity and Systemic Activities of Fungicidal Triazoles and Bleaching Herbicidal Compounds

Shizuya Tanaka, Masahiro Takahashi, Yuji Funaki[1], Kazuo Izumi[2],
Hirotaka Takano, and Masakazu Miyakado

Agricultural Chemicals Research Laboratory, Sumitomo Chemical
Company, Ltd., 4–2–1 Takatsukasa, Takarazuka, Hyogo 665, Japan

Two examples are described to show that the systemic translocation as well as the systemic activity is markedly lowered when the log *P* value of series of analogs exceeds a boundary located between 4 and 5. Triazole fungicides such as uniconazole, diniconazole, and their analogs were active against powdery mildew on barley by foliage application. By soil application, however, the fungicidal activity depended on whether the log *P* value of the compounds is lower or higher than 4. Herbicidal compounds with bleaching activity such as 3-phenoxybenzamides, fluridone and its analogs, norflurazon, difunon, flurtamone, and diflufenican showed three types of bleaching patterns on expanded cotyledons of radish after seed treatment. In spite of their structural diversity, variations in the bleaching pattern was clearly understandable in relation to the log *P* value of the compounds. When the log *P* value of the compounds was between 4 and 5, pronounced changes in the systemic activities occurring through the root-to-shoot as well as the lateral and translaminal translocations were observed.

The systemic translocation of agrochemicals in plants is a very important problem. For agrochemicals applied at remote locations from the targeted zone in plants, the systemicity is a prerequisite for their bioactivity. Although the translocation of agrochemicals in plants has been shown to be connected to their hydrophobicity in terms of the log *P* value (*P*: 1-octanol / water partition coefficient) in certain respects (*1*, *2*, *3*, *4*, *5*), "the systemic activity" of agrochemicals in relation to their systemicity in plants has not always been well understood (*3*). In this chapter, we selected two topics dealing with fungicidal activity of triazole compounds and bleaching activity of

---

[1] Current address: Technology & Research Management Office, Sumitomo Chemical Company, Ltd., 2–27–1 Shinkawa, Chuo-ku, Tokyo 104, Japan
[2] Current address: Farmchemicals & Materials Division, Sumitomo Chemical Company, Ltd., 4–5–33 Kitahama, Chuo-ku, Osaka 541, Japan

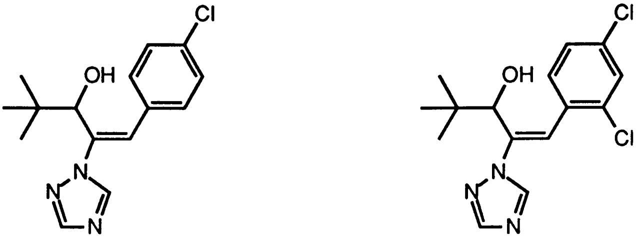
0097–6156/95/0606–0107$12.00/0

herbicidal compounds. In both cases, systemic activity of agrochemicals is controlled greatly by their log *P* value. There seems to exist a boundary in the log *P* value above which the systemicity decreases drastically so that the systemic activity almost disappears under certain conditions.

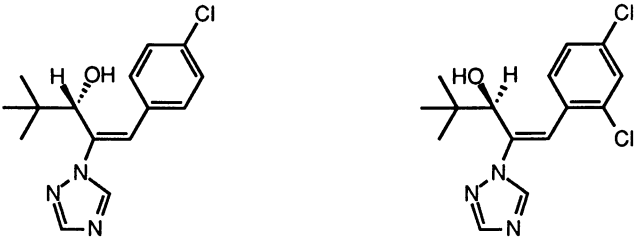
Fungicidal Triazole Compounds

Triazole compounds, uniconazole and diniconazole (Figure 1), are not only fungicidal but also exhibit a plant growth regulatory activity (PGR activity) as their racemic form. During the course of their biological evaluation as agrochemicals, we observed a large difference in systemic activity between these two compounds. Limited systemicity of diniconazole as a candidate of novel fungicide was an important factor to reduce and separate the PGR activity. On the other hand, the higher xylem-mobility of uniconazole was essential as the PGR for its use as granular formulations in the paddy field. After extensive toxicological and environmental studies, two optically active compounds, uniconazole P (the *S*–isomer of uniconazole) as a PGR, and diniconazole M (the *R*–isomer of diniconazole) as an agricultural fungicide, were eventually marketed.

Both optically active triazoles have been manufactured industrially by enantioselective reduction of their corresponding ketone precursors (*6*). Resolution into the optical isomers has resulted in a reduction of the application rate and an enhanced selectivity between target and non-target organisms. Moreover, an almost complete separation of fungicidal and PGR activities has been accomplished, *i.e.*, uniconazole P shows no significant fungicidal activity and diniconazole M has almost no PGR effect (*7*). In the following sections, however, all experimental results were obtained using racemic compounds.



uniconazole **diniconazole**



uniconazole P [*S*(+)-isomer]
use: plant growth regulator
inhibition: gibberellin biosynthesis

diniconazole M [*R*(-)-isomer]
use: fungicide
inhibition: ergosterol biosynthesis

Figure 1. Uniconazole, Diniconazole, Uniconazole P, and Diniconazole M.

Translocation of Uniconazole in Cucumber: Cucumber treated with radiolabeled uniconazole (log *P*: 3.77) at the roots is shown in Figure 2. It is suggested that uniconazole is highly xylem-mobile from roots. Within 2h, it moves into veins of the first and second foliage leaves. It spreads over all of the parts of these foliage leaves and cotyledons after 6h, and reaches to the third foliage leaf 24h after the start of the treatment.

When a central area on the second foliage leaf was treated by uniconazole, on the other hand, the translocation through xylem across the cuticular membrane is not observed clearly, but it is laterally mobile and maybe translaminally mobile as well to a lower extent as shown in Figure 3. After 2h, the xylem-translocation toward the leaf tip occurs first, but a lateral translocation is observed only after 6, 24, and 48h. The radioactive area expands about 4 fold during the 48h treatment. The lateral translocation on the leaf is much slower than the root-to-shoot translocation. This is probably due to the fact that, once uniconazole is trapped in the cuticular membrane from the leaf surface, it is slow to diffuse out and translocate again. Such hydrophobic agrochemicals as PCNB (log *P*: 5.0) and fluvalinate (log *P*: 5.5) are reported to be incorporated into cucumber leaf cuticles isolated with enzymatic treatments and to reach saturation rapidly, but have difficulty getting across the cuticular membrane (*8*). Translocation through phloem to other leaves and roots of uniconazole was not observed.

Systemic Activity of Triazole Compounds: Protective activity of uni(dini)conazole analogs against the powdery mildew on barley are shown in Table I. Uniconazole and compounds **1** and **2** were highly active on foliage as well as soil application. As

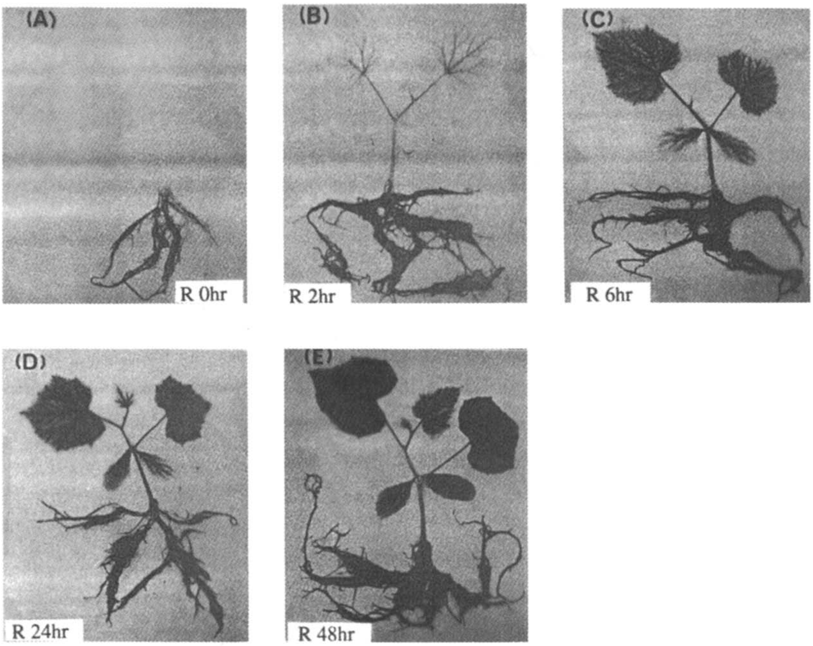


Figure 2. Translocation of ^{14}C-Uniconazole in Cucumber by the Root Treatment. (A): 0h. (B): 2h. (C): 6h. (D): 24h. (E): 48h.

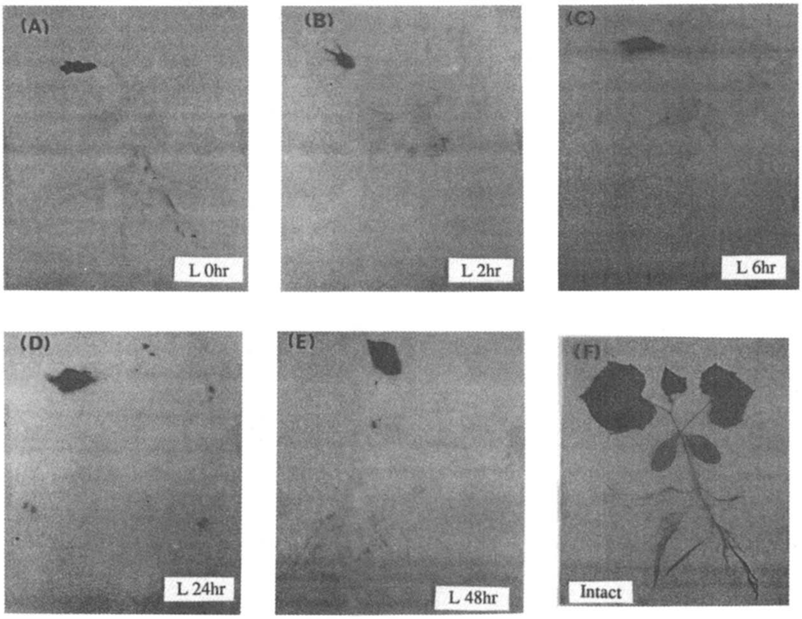


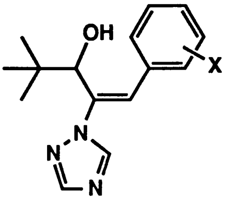
Figure 3. Translocation of ^{14}C-Uniconazole in Cucumber by the Foliage Treatment.
(A): 0h. (B): 2h. (C): 6h. (D): 24h. (E): 48h.
(F): Photograph of cucumber used in this experiment.

indicated above, the translocation of uniconazole is very slow after the foliage treatment, but this is when the application has been made topically on leaves. With the spray on the entire foliage, uniconazole was believed to be distributed in plants much faster and the difference between soil and foliage treatments does not seem to be significant.

Among the four compounds, diniconazole (log *P*: 4.30) was not active by soil application. Because of the highest log *P* value, the soil adsorption of diniconazole is also highest (*9*). Under conditions in which the soil adsorption is not critical, however, we estimated that its root-to-shoot translocation is still limited. The application of diniconazole on roots of barley seedlings did not show evidence for its significant mobility. The PGR activity by soil application is also shown in Table I. Diniconazole exhibited only a very low PGR activity compared with the other three compounds. Table I shows that the systemic activity through the root-to-shoot translocation of these triazole compounds decreases drastically when their log *P* value exceeds 4.

Systemic Activity and Hydrophobicity: To examine the relationship between the log *P* value and the systemic activity of uni(dini)conazole analogs further, we prepared a number of derivatives. For those in which the 4-Cl substituent on the benzene ring of uniconazole was replaced by a variety of less hydrophobic substituents, the protective activities against the powdery mildew on barley and the brown rust on wheat were measured by the foliage application (*10*) as summarized in

Table I. Fungicidal and PGR Activities of Triazole Compounds



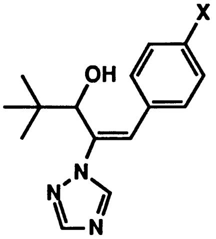
			Powdery Mildew on Barley[a]						PGR[b]
			foliage			soil			soil
Compd	X	log *P*	50	12.5 (mg/l)	3.1	5	1.25 (mg/pot)	0.3	5 (g/a)
unic[c]	4-Cl	3.77[d]	4	4	3	4	4	4	4
1	2-F,4-Cl	3.91[e]	4	4	4	4	4	4	3
2	4-Br	3.92[e]	4	3	2	4	3	2	3
dinic[f]	2,4-Cl_2	4.30[d]	4	4	4	1	0	0	0

[a] Disease control, 0: <50%, 1: 50–74%, 2: 75–89%, 3: 90–99%, 4: 100% control of infection. [b] Growth suppression of the rice plant *vs*. control, 0: <30%, 1: 31–50%, 2: 51–70%, 3: 71–90%, 4: >90%. [c] Uniconazole. [d] Newly measured value by the shake-flask method. [e] Calculated additively with the CLOGP program (MedChem Software Release 3.55, BioByte Corporation, Pomona College, Claremont, California, 1994) on the basis of the experimental value of uniconazole. [f] Diniconazole.

Table II. With the substituent modifications, the intrinsic fungicidal potency should be varied. Although the fungicidal effect was lower than uniconazole in the less hydrophobic analogs **3–10,** compounds having electron-withdrawing substituents, especially the 4-CN compound **7,** show a moderate activity (*10*). This suggests that there are possibilities to find potent compounds with lower hydrophobicity than uniconazole.

In fact, of compounds **11** and **12** in which the *t*-butyl-methylol side chain moiety of uniconazole is modified, the compound **11** less hydrophobic is as highly active as uniconazole by soil application as shown in Table III. Diniconazole and compound **12**, whose log *P* values are almost the same, were active by the foliage application, but inactive by the soil application. The xylem mobility of these two compounds should be quite restricted. Compound **11** was less active than uniconazole by foliage application, indicating that the relative significance of the root-to-shoot translocation, is highest in compound **11** among the four in Table III. The above results indicate that the activity measured with the soil application reflects the root-to-shoot systemicity in barley, and that the higher the hydrophobicity within a range between 2.0 and 4.5 in terms of log *P*, the lower is the systemicity. More important is the fact that there is a boundary in the log *P* value between compound **2** (log *P*: 3.92) and diniconazole (log *P*: 4.30) above which an abrupt decrease in the systemic activity is observed.

Table II. Fungicidal Activities of Uniconazole and Its Analogs[a]



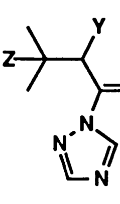
			Barley powdery mildew[c]			Wheat brown rust[c]		
Compd	X	log P[b]	50	12.5 (mg/l)	3.1	50	12.5 (mg/l)	3.1
3	$SOCH_3$	1.47	4	3	1	0	0	0
4	NH_2	1.83	0	0	0	0	-[d]	-
5	$NHCOCH_3$	2.08	0	0	0	0	-	-
6	OH	2.39	0	0	0	0	-	-
7	CN	2.49	4	3	1	3	2	1
8	NO_2	2.80	4	2	1	0	0	0
9	OCH_3	2.98	2	0	0	0	-	-
10	$COOC_2H_5$	3.56	0	0	0	0	-	-
unic[e]	Cl	3.77[f]	4	4	3	4	4	4

[a] Adapted from ref. 10. [b] Calculated additively with the CLOGP program on the basis of the value of uniconazole. [c] Disease control, 0: <50%, 1: 50–74%, 2: 75–89%, 3: 90–99%, 4: 100% control of infection. [d] Not detected. [e] Uniconazole. [f] Experimentally measured value by the shake-flask method.

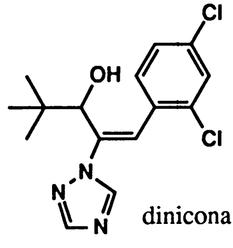
The above indications are important in designing better agrochemicals, the log P value of which is close to 4.0. As mentioned above, compound **11** is xylem-mobile effective. Nevertheless, it should be possible to make its log P value (2.02) closer to that of uniconazole (3.77) to improve the intrinsic activity without losing the systemic activity significantly, perhaps by introducing another Cl substituent at the 2–position on the benzene ring. The introduction of substituents into the molecule should be made so that the molecular log P value should not exceed about 4 to maintain the systemic activity through the root-to-shoot translocation.

As mentioned in the introductory section, the root-to-shoot translocation has been examined in connection with the log P value of agrochemicals. Briggs *et al.* (*1*) have shown that the translocation of neutral compounds in terms of the ratio of the equilibrated concentration in the barley xylem sap to that in the external solution in which roots are immersed is related with their log P value in the manner of a bell-shaped curve, the optimum log P value being 1.8. The most desirable root-to-shoot translocation property is obtained in compounds the log P value of which is between

Table III. Fungicidal Activity of Triazole Compounds by the Foliage and Soil Treatment



uniconazole Y=OH, Z=CH_3
11 Y=OH, Z=OH
12 Y=OCH_3, Z=CH_3



diniconazole

		Powdery mildew on Barley[b]					
		foliage			soil		
Compd	log *P*[a]	50	12.5 (mg/l)	3.1	5	1.2 (mg/pot)	0.3
11	2.02	3	1	0	4	4	4
unic[c]	3.77	4	4	3	4	4	4
dinic[d]	4.30	4	4	4	1	0	0
12	4.39	4	3	2	0	0	0

[a] Calculated additively with the CLOGP program on the basis of the value of uniconazole. [b] Disease control, 0: <50%, 1: 50–74%, 2: 75 –89 %, 3: 90 – 99 %, 4: 100 % control of infection. [c] Uniconazole. [d] Diniconazole.

0.5 and 3.0. They have also emphasized their observation that the reported behavior of a number of agrochemicals in various plant species conforms with their mathematical formulations.

Hsu *et al.* (*2*) have also examined the concentration ratio between the soybean xylem sap and the external solution using different types of compounds under different experimental conditions from those of Briggs *et al.* Their overall results are similar to those of Briggs *et al.*, except that the optimum log *P* value is about 3 and the favorable log *P* range for the systemicity is between 2.5 and 4.5. The relationship of the root-to-shoot translocation with hydrophobicity seems to be dependent upon plant species, structural series of agrochemicals, and experimental conditions. It could be said, however, that both of the optimum log *P* values for the root-to-shoot translocation from two independent research groups are lower than the log *P* value of uniconazole. The sequence of the xylem mobility of uni(dini)conazole analogs estimated from the systemic activity by the soil application shown in Tables I to III, seems to be consistent with the bell-shaped relationship observed by Briggs *et al.* and Hsu *et al.*

Distribution Patterns of Fungicides and Herbicides According to Log *P*: Apart from discussions based on experimental results, there is a kind of "evidence" showing the importance of the root-to-shoot systemicity in the activity of fungicides and

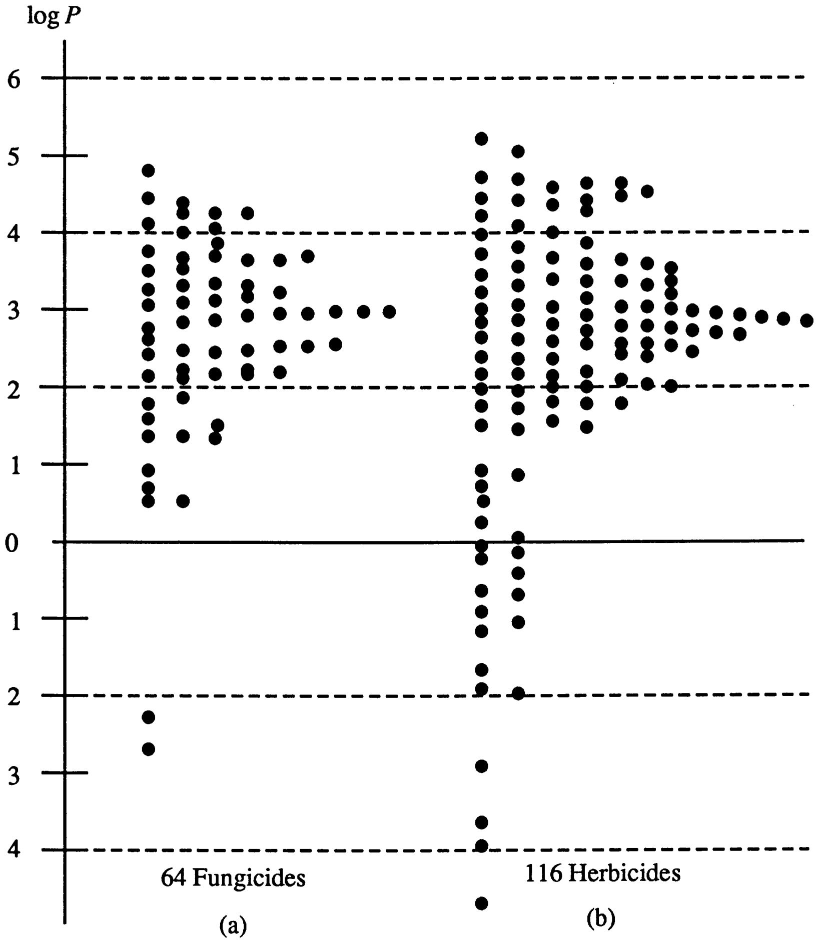


Figure 4. Distribution Patterns of Fungicides and Herbicides according to Log *P*.

herbicides. Figures 4a and 4b are histograms of fungicides and herbicides respectively, plotted against the log *P* value. In these histograms, only agrochemicals whose log *P* value is available in the Pesticide Manual (*11*) were picked up. For 64 fungicides (F) and 116 herbicides (H), the distribution pattern is very similar. The maximum population is located at about 3 in terms of log *P* with a shoulder at about 4.5. In hydrophilic compounds whose log *P* value is below zero, sulfonylurea herbicides, organophosphorus compounds such as glyphosate (H) and fosetyl aluminium (F), and cymoxanyl (F) are included. The upper limit in the log *P* value in fungicides as well as herbicides is about 5. There is no fungicide and herbicide whose log *P* value exceeds 5, except for two soil-applied dinitroaniline herbicides, pendimethalin (log *P*: 5.18) and trifluralin (log *P*: 5.34), which are known to be taken up through the vapor moisture from soil but not through xylem (*12*).

It is interesting to note that these bell-shaped histograms look very similar to the plot of the concentration ratio between xylem sap and external solution against the log *P* value of a number of neutral compounds. The histograms are especially similar to the plot obtained by Hsu *et al.*(*2*), in that the "optimum" is located at about 3. Although the mechanisms of uptake into and translocation inside plants are not necessarily uniform, the similarity between the population and the xylem-mobility index, when they are plotted against the log *P* value, would suggest that the root-to-shoot mobility is one of the most important mechanisms of the translocation in plants for presently available agrochemicals.

Herbicidal Compounds with Bleaching Activity

Carotenoids in plants protect the photosynthetic apparatus from destruction by quenching the active singlet oxygen atoms produced in the photosynthetic process. Compounds that interfere with carotenoid biosynthesis induce destruction of chlorophylls and cause bleaching of plants. A number of herbicidal compounds with bleaching activity are known as inhibitors of carotenoid biosynthesis, and some of them are shown in Figure 5.

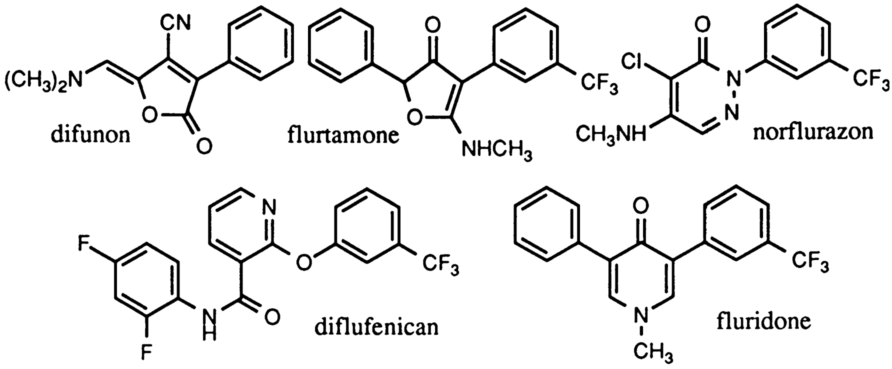
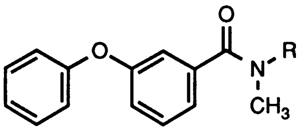


Figure 5. Chemical Structure of Bleaching Herbicides.

Bleaching Patterns in Radish Seedlings: 3-Phenoxybenzamides **13**, **14**, and **15** (Table IV) are a series of compounds having herbicidal bleaching activity (*13*). These *N*-methyl-*N*-higher alkyl-type homologs were prepared in order to bring the log *P* value within the range between 3.5 and 5. Radish seeds soaked in the solution containing these compounds (1.6 mM) for 24h before germination were sown in pots and grown, and the expanded cotyledons were observed after 5 days. They showed three different bleaching patterns as shown in Figure 6 and Table IV. Compound **13** gave pattern A in which the whole cotyledons are bleached. Compound **14** induced pattern B in which the large cotyledon is bleached entirely, but the small one is bleached only at the margins. Compound **15** exhibited pattern C in which the margins of the adaxial surfaces of the both cotyledons and the abaxial surface of the small cotyledon are bleached, but the abaxial surface of the large cotyledon is bleached entirely.

Radish seeds were treated with three radiolabeled insecticidal compounds, metolcarb (log *P*: 1.70), fenitrothion (log *P*: 3.43), and fenpropathrin (log *P*: 6.00), similar to the above herbicidal compounds. The less hydrophobic compounds, metolcarb and fenitrothion, showed a complete distribution of radioactivity on the expanded cotyledons, but fenpropathrin showed only a marginal distribution. From the behaviors of these radiolabeled insecticides, it is supposed that compounds **13**, **14**, Table IV. 3-Phenoxybenzamides and Their Bleaching Patterns[a]



				Leaf surface[d]			
Compd	R	log P[b]	Bleaching pattern[c]	1	2	3	4
13	n-C_4H_9	3.89	A	W[e]	W	W	W
14	n-C_5H_{11}	4.43	B	W	W	M[f]	M
15	n-C_6H_{13}	5.01[g]	C	W	M	M	M

[a] Adapted from ref. 13. [b] Newly measured value by the shake-flask method [c] See Figure 6. [d] 1: Abaxial surface of the large cotyledon, 2: Adaxial surface of the large cotyledon, 3: Adaxial surface of the small cotyledon, 4: Abaxial surface of the small cotyledon. [e] Whole surface was bleached. [f] Only marginal area was bleached. [g] Estimated value by the HPLC method.

and **15** were actually absorbed into seeds and then translocated to the bleached area of the cotyledons. When roots of etiolated seedlings were treated with compounds **13**, **14**, and **15**, the bleaching was observed on hypocotyl and marginal areas of both cotyledons after 48h. However, the bleaching patterns on the cotyledons by the root treatment were different from any of the patterns A, B, and C of the seed treatment.

In radish seeds, the small cotyledon is covered by the large one, and surfaces of both cotyledons are folded tightly in the seed coat in the following order from the outer side as shown in Figure 7: the abaxial surface (surface 1) of the large cotyledon, its adaxial surface (surface 2), the adaxial surface (surface 3) of the small one, and

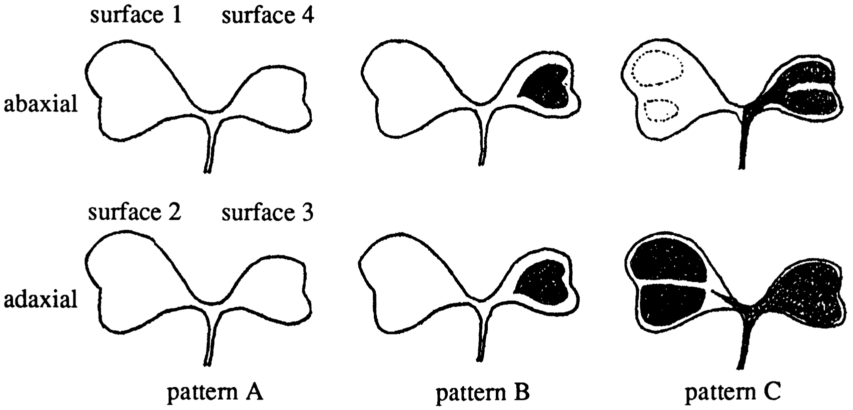


Figure 6. Bleaching Patterns on Radish Cotyledons.

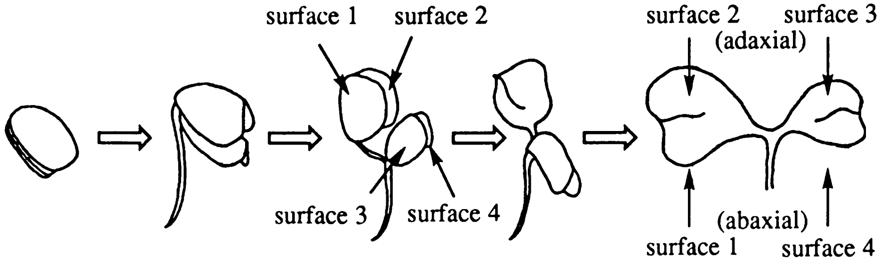
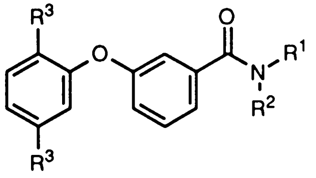


Figure 7. Development of Radish Cotyledons.

its abaxial surface (surface 4). The test compounds were first taken up through the seed coat at the whole area of surface 1 and the marginal areas of the other three surfaces. Then, they were translocated laterally from margins to the central areas and translaminally from the outer surface 1 to the inner surfaces 2–4, to cause three different bleaching patterns depending on their mobilities. It is considered that the lateral and translaminal mobilities of compound **13** showing the pattern A are high, and those of compound **14** showing the pattern B are moderate, and those of compound **15** showing the pattern C are very limited.

Bleaching Patterns and Hydrophobicity: The bleaching patterns caused by the *N*-monoalkyl-3-phenoxy benzamides, **16**, **17**, and **18** are compared with the patterns induced by the above-mentioned *N*,*N*-dialkyl homologs **13**, **14**, and **15**. As shown in Table V, each pair of compounds having nearly equal log *P* values exhibited very

Table V. Bleaching Patterns of 3-Phenoxybenzamides[a]



Compd	R^1	R^2	R^3	log *P*[b]	Bleaching Pattern[c]
16	CH_3	H	CH_3	3.76	A
13	*n*-C_4H_9	CH_3	H	3.89	A
17	*i*-C_3H_7	H	CH_3	4.35	B
14	*n*-C_5H_{11}	CH_3	H	4.43	B
18	*t*-C_4H_9	H	CH_3	4.78	C
15	*n*-C_6H_{13}	CH_3	H	5.01[d]	C

[a] Adapted from ref. 13. [b] Newly measured values by the shake-flask method except compound **15**. [c] See Figure 6. [d] Estimated value by the HPLC method.

Table VI. Bleaching Herbicides and Their Bleaching Patterns[a]

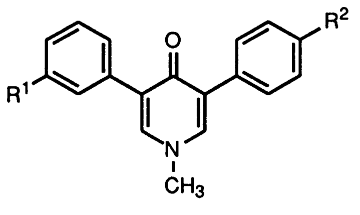
Compd	log *P*	Bleaching pattern[b]
difunon	1.07	A
norflurazon	2.52	A
flurtamone	3.22	A
diflufenican	4.49	C

[a] Adapted from ref. 13. [b] See Figure 6.

similar bleaching patterns. The log *P* boundary discriminating compounds showing the patterns A and B is located in the range of 3.9 to 4.4, and that differentiating compounds exhibiting the patterns B and C is in the range of 4.4 to 4.8. The lateral and translaminal mobilities vary from high to moderate within the first log *P* range, and then from moderate to limited within the second range.

The activity of such known bleaching herbicides (Figure 5) as difunon, norflurazon, flurtamone, and diflufenican, and fluridone and its analogs **19**, **20**, and **21**, were also examined as shown in Tables VI and VII (*14*). They showed either pattern A or C depending on their log *P* values. Pattern B was not observed. From the data in Tables V–VII, the lowest log *P* values showing pattern C seem to be located between 4.5 and 5.3. The log *P* value (4.59) of compound **20** showing pattern A is slightly greater than that of diflufenican (4.49) showing pattern C. It is emphasized, however, that the systemic activity through lateral and translaminal translocation of herbicidal compounds is also drastically lowered above the log *P* values in the range of 4.5 to 5.3 depending upon the structural features of compounds as far as the radish seed test is concerned.

Table VII. Bleaching Patterns of Fluridone and Its Analogs[a]



Compd	R[1]	R[2]	log *P*	Bleaching pattern[b]
fluridone	CF_3	H	2.98[c]	A
19	*n*-$C_5H_{11}O$	H	4.07[d]	A
20	*n*-$C_6H_{13}O$	H	4.59[d]	A
21	*n*-$C_6H_{13}O$	Cl	5.23[e]	C

[a] Adapted from ref. 14. [b] See Figure 6. [c] Taken from ref. 14.
[d] Newly measured values by the shake-flask method.
[e] Estimated value by the HPLC method.

Conclusion

As described in above examples, a drastic change in the systemic activity of fungicidal and herbicidal compounds occurs when their log *P* values are in the range of 4 to 5. The upper limit of the log *P* value of known fungicides and herbicides is slightly above 5. Great care must be taken for structural modification in this log *P* range to control the systemic activity. The lower limit, if any, may be another possible research target.

Acknowledgments

We appreciate our colleagues, Drs. Kenzo Ueda, Hiromichi Oshio, and Chiyozo Takayama, for their valuable discussions and suggestions. We also acknowledge a number of people of our company involved in the development of uniconazole P and diniconazole M.

Literature Cited

1. Briggs, G. G.; Bromilow, R. H.; Evans, A. A. *Pestic. Sci.* **1982,** *13,* 495.
2. Hsu, F. C.; Marxmiller, R. L.; Yang, A. Y. S. *Plant Physiol.* **1990,** *93,* 1573.
3. Uchida, M. *Pestic. Biochem. Physiol.* **1980,** *14,* 249.
4. Uchida, M.; Nishizawa, H.; Suzuki, T. *J. Pestic. Sci.* **1982,** *7,* 397.
5. Kleier, D. A. *Pestic. Sci.* **1994,** *42,* 1.
6. Tanaka, S. ; Funaki, Y.; Suzukamo, G.; Yoneyoshi, Y.; Tada, K. *J. Chem. Soc. Jpn., Chem. Indust. Chem.* **1994,** *11,* 953.
7. Takano, H.; Inoue, S.; Oshio, H.; Kamoshita, K.; Kobayashi, K.; Ogoshi, A. *Ann. Phytopath. Soc. Jpn.* **1992,** *58,* 691.
8. Ohori, Y.; Ihashi, Y. *Mitsubishi Chemical R & D Review* **1987,** *1,* 22.
9. Uchida, M.; Kasai, T. *J. Pestic. Sci.* **1980,** *5,* 553.
10. Funaki, Y.; Ishiguri, Y.; Kato, T.; Tanaka, S. *J. Pestic. Sci.* **1984,** *9,* 229.
11. *The Pesticide Manual*; Worthing, C. R.; Hance, R. J., Eds.; 9th Ed.; British Crop Protection Council: Farnham, Surrey, 1991.
12. Asthenia, F. M.; Craft, S. S. *Mode of Action of Herbicides*; 2nd Ed.; John Wiley & Sons: New York, 1981; pp 208–210.
13. Takahashi, M.; Miyakado, M.; Tanaka, S. *J. Pestic. Sci.* **1993,** *18* , 169.
14. Takahashi, M.; Kawamura, S.; Miyakado, M.; Sanemitsu, Y.; Tanaka, S. *Pestic. Sci.* **1993,** *39,* 169.

RECEIVED April 26, 1995

Chapter 9

The Incidence of Delayed-Type Contact Dermatitis in Modern Pesticides

Philip S. Magee

BIOSAR Research Project, 141 Sealion Place, Vallejo, CA 94591 and School of Medicine, University of California, San Francisco, CA 94143

A recently developed model for allergic contact dermatitis (ACD) is applied to 148 pesticides to assess both allergens and nonallergens. Those reported to be moderate to strong irritants, but scoring as probable non-allergens, are classified as immediate type primary irritants. Although there is no certainty in the absence of animal testing, the performance of the model with respect to non-irritating pesticides is exceptional and those classified as ACD are considered as highly probable allergens based on previous validation experience. To the credit of the modern pesticide industry, only 13 of the 148 pesticides were identified as probable ACD problems with 25 others showing some form of direct irritation. As the seriousness of dermal irritation in the handling of pesticides becomes more clearly recognized, it can be expected that early animal testing will allow few of these problems to reach the future market without appropriate handling precautions.

Both acute and chronic hazards of pesticides have long been recognized by obligatory animal testing to cover every reasonable toxicology problem that plant personnel, shippers, farmers and the ingesting public might encounter. Thus, it is common to find the oral and dermal LD_{50} against common laboratory animals and often some mention of chronic feeding results in most pesticide compilations. Less commonly reported are the non-lethal effects of lung and skin irritation which can lead to serious job and personal problems. While an exposed worker can easily protect from lung irritation when the hazard is known, equivalent protection from skin irritation is much more difficult and virtually impossible for field workers in the heat of summer. The ultimate solution, of course, is the denial of registration for future pesticides exhibiting severe lung and dermal irritation. The purpose of this study is to assess the nature of dermal irritation of currently registered modern pesticides. To accomplish this goal, a large sample of pesticides was selected from The Agrochemicals Handbook(*1*), with the following distribution:

0097–6156/95/0606–0120$12.00/0

Class	Number
insecticides	27
herbicides	61
fungicides	42
other	18 (miticides, ovicides, growth regulators, etc.)
TOTAL	148

In all of the selected cases, there was sufficient mention of dermal effects to allow a numerical rating of severity ranging from 1.0 to 0.0. Handbook comments are rated as follows:

Comment	Rating
strong skin irritant	1.0
inflammation of skin	1.0
allergen/sensitizer	1.0
moderate skin irritation	0.7
local skin irritation	0.7
possible skin irritant	0.5
no mention of irritation	0.5
slight skin irritation	0.3
no skin irritation	0.0

Note that the ambiguous statements are given an arbitrary nul or indeterminate rating of 0.5. Only 9 of the 148 pesticides were so rated and do not detract from the firmer ratings of 0.0-0.3 and 0.7-1.0 which provide the basis for this study. By comparing these experiential ratings with the results of the classification model, we hope to achieve a more incisive look into pesticide dermal problems. The model detects compounds likely to induce allergic contact dermatitis or delayed type hypersensitivity, a T-cell mediated immune response reacting in 24-72 hours after initial sensitization(*2*). Immediate type hypersensitivity (contact urticaria or hives) is a different immune response based on rapid humoral reaction with immunoglobin-E in sensitive individuals. In addition, there are several other irritation responses that do not involve the immune system.. As examples of each type, bromoxynil is a strong immediate type irritant, especially of the mucous membranes (nasal, lung), while captan, a protein reactive, neutral sulfenylating agent, causes delayed contact dermatitis (ACD). Some chemicals, such as 2,4-dinitrochlorobenzene and phthalic anhydride, are capable of both immediate and delayed responses.

Immune System Activation

Activation of the immune system as explained in detail by von Blomberg, Bruynzeel and Scheper (*2*) can be summarized as follows. Penetration of the stratum corneum into the viable epidermis is required to reach the domain of the allergen presenting Langerhans cells. The causative, protein-reactive allergen may be directly reactive (hapten) or may require prior metabolic activation in the epidermal tissue (prohapten).

In either case, the activating reaction occurs with surface molecules on the Langerhans cells. Reactions may occur with reactive residues such as cysteine, lysine, histidine and serine (nucleophilic), cystine and carboxamide linkages (electrophilic) or tyrosine (radical). The most common hapten reactants are electrophilic (anhydrides, activated halides, conjugated olefins, aldehydes, epoxides, quinones), while strong nucleophiles such as primary amines and thiols form a smaller class(*3*). The incidence of irritant contact dermatitis from the nicotine smoking withdrawal patch can be as high as 30% (*4*), a relevant point as free nicotine is still used as an insecticidal fumigant and the sulfate for minor insecticidal uses(*1*). Radical reactions with cell surface molecules necessarily require metabolic activation *v'a* oxidative or dehalogenating enzymes. DDT has been implicated in this type of reactivity (*5*).

Following reaction of the hapten with the Langerhans cell surface molecules, the allergen presenting cell migrates by an afferent route to the paracortical region of the local lymph node where reversible non-covalent encounters with approximately 100,000,000 variations of helper T-cells take place. The receptor sites on the T-cells are too large to recognize individual haptens, but are capable through reversible trial and error of achieving a best fit (lowest free energy) with the hapten-modified cell surface molecules. Completion of this dynamic process triggers the clonal proliferation of best- fit T-cell lymphocytes, which eventually leave the lymph node by an efferent route (thoracic duct) to reenter the body and the viable epidermis. This simplified description represents the sensitizing stage of allergic contact dermatitis. The clonal proliferation of the T-cells can be quantitatively measured by the murine local lymph node assay (LLNA) of Kimber and Weisenberger (*6*), a method well suited to QSAR research (*7*).

A human or animal subject may remain sensitized for months or even years before a second challenge of the same allergen (hapten or prohapten) occurs. The violent eruptive nature of the allergic response generally occurs in 24-72 hours after contact, hence the description: delayed-type hypersensitivity. The most universally recognized prohapten is urushiol (poison oak, ivy), a mixture of catechols requiring metabolic activation to the protein reactive quinone. Urushiol has been identified in Chinese lacquers used for over 2000 years (*8*).

Modeling Allergic Contact Dermatitis

One of the difficulties in developing a QSAR model for ACD is selection of reliable data. The confusion is illustrated by twenty literature entries for formaldehyde in a private database (*9*). Despite the unquestioned protein reactivity of formaldehyde, different investigators found responses ranging from mild to strong. For this reason, the model contains only allergens certified to be of moderate to strong ACD response and covering a broad range of fundamental organic structure (n = 36). A matching set of non-allergens covering a similar range of structure was selected from cosmetic and fragrance compilations (*10,11*). The selection procedure is detailed in a recent reference (*12*).

Tables 1-3 contain the selected sets of strong, moderate and miscellaneous allergens used in this study. Tables 1 and 2 contain data drawn from the Ziegler database (*9*) rated 3 and 2 respectively on a scale of 0-3. Table 3 contains data from

Dupuis and Benezra (*3*), from the local lymph node assay (*6*) and from the clinical experience of Maibach (*12*). Table 2 contains the only outlier observed in the correlation (geraniol). This was anticipated as geraniol is the trans isomer of nerol, an established aroma chemical (*10*) and both have the same descriptors. Nerol is correctly classified as non-allergenic while geraniol is incorrectly classified and was removed from the equation as an outlier. A laboratory study is in progress to determine if nerol and geraniol metabolize differently in the viable epidermis. The tables contain the calculated values for MR, PL, HBA and HBD, as described immediately after the tables. Coding for the dichotomous descriptors representing hapten and prohapten activity is omitted as self-evident from simple inspection of the molecular structure (description follows).

Table 1. Severe Allergens *versus* Non-Allergens.

Case	Compound	Class	MR	PL	HBA	HBD
1	Butanediol Diacrylate	1	52.26	4.54	8	0
2	Cinnamaldehyde	1	42.20	2.72	2	0
3	Dihydrocoumarin	1	39.47	2.15	4	0
4	2,4-Dinitro-1-chlorobenzene	1	44.05	2.26	4	0
5	Hydroquinone	1	20.03	1.67	4	2
6	Isoeugenol	1	49.23	3.46	4	1
7	Kelthane	1	86.78	6.68	2	1
8	a,a-Methyl Cinnamaldehyde	1	46.82	3.13	2	0
9	p-Phenylenediamine	1	35.17	1.67	2	4
10	N-Phenyl N-a-naphthylamine	1	70.52	4.29	1	1
11	3,5,3',4'-Tetrachlorosalicylanilide	1	81.32	6.41	5	2
12	4-Chloro-m-cresol	1	37.83	3.54	2	1
13	Indole	0	38.68	2.79	0	1
14	a-Dimethyl Phenylethylacetate	0	55.79	5.06	4	0
15	3,5-Dimethyl-5-formyl-cyclohexene	0	41.47	3.71	2	0
16	1-Bromo-2-phenylethylene	0	44.20	3.36	0	0
17	6-Isopropyl Quinoline	0	54.62	4.76	1	0
18	4-Methoxybenzyl Alcohol	0	39.27	3.05	4	1
19	Beta-naphthyl Methyl Ketone	0	52.27	3.84	2	0
20	2-Phenylethanol	0	37.45	3.00	2	1
21	Gamma-nonalactone	0	42.86	4.32	4	0
22	Methyl 2-Nonenoate	0	50.26	5.26	4	0
23	Methyl 2-Aminobenzoate	0	41.30	2.44	5	2
24	Linalool	0	51.52	5.05	2	1

Table 2. Moderate allergens *versus* non-allergens.

Case	Compound	Class	MR	PL	HBA	HBD
1	Ethylenethiourea	1	30.40	1.54	0	2
2	Ethylene Glycol Dimethacrylate	1	52.26	4.13	8	0
3	Eugenol	1	49.23	3.35	4	1
4	Geraniol (*trans* isomer of Nerol)	1	51.52	5.18	2	1
5	Glycidyl Benzoate	1	45.91	3.46	6	0
6	Hexachlorophene	1	88.98	8.42	4	2
7	1,2,6-Hexanetriol	1	35.24	3.13	6	3
8	Methyl 4-Hydroxybenzoate	1	38.73	2.56	6	1
9	4-Hydroxybenzoic Acid	1	34.11	1.67	6	2
10	2-Hydroxyethyl Methacrylate	1	33.60	2.79	6	1
11	Indomethacin	1	92.66	6.40	8	1
12	4-Isopropylaminodiphenylamine	1	73.36	5.29	4	2
13	2-Methoxy-4-propenylphenyl Acetate	0	57.93	3.32	6	0
14	4-Methoxyacetophenone	0	43.30	3.44	4	0
15	Nerol (*cis* isomer of Geraniol)	0	51.52	5.18	2	1
16	5-Methyl-3-heptanone Oxime	0	43.65	4.23	3	1
17	4-Methoxybenzaldehyde	0	38.68	2.79	4	0
18	10-Undecen-1-ol	0	54.83	5.21	2	0
19	Gamma-Undecalactone	0	52.10	5.40	4	0
20	a,a-Dimethyl Phenethyl Alcohol	0	46.69	3.73	2	1
21	2-Isobutyl Quinoline	0	57.95	5.19	1	0
22	Diethyl Phthalate	0	56.67	3.30	8	0
23	Serine	0	23.41	0.85	7	4
24	Tyrosine	0	47.74	2.52	7	4

and dichotomous. Continuous descriptors such as MR, LogP(octanol/water) and its factors (PL and PH)(*13*) were used to model transport and reversible binding. Molar Refraction (MR) models the induced dipole moments of London Forces, while the factors PL and PH are the contributions of the lipophilic and hydrophilic substructures in the calculation of LogP. Ordinal descriptors were used to describe the propensity of the molecules to form hydrogen bonds with proteins encountered in the stratum corneum and viable epidermis. The electron pair count on oxygen and nitrogen provides a descriptor for potential acceptor bonds (HBA), while the counts of N-H and O-H groups describe potential donor bonds (HBD). An exception is the nitro group where both oxygen atoms are delocalized and the nitrogen atom is nonbasic. Only one available acceptor pair is considered for each oxygen, giving an HBA value of 2 for each nitro group. By contrast, both electron pairs are considered in a carbonyl group. Finally, groups with potential protein reactivity as haptens or metabolizable as prohaptens were coded as indicator variables (1.0/0.0). These are test variables that respond statistically when reactive potential is realized. With few exceptions, such as strong nucleophiles (nicotine, ethylenediamine), hapten reactivity is expressed by the uncatalyzed reaction of electrophiles (aldehydes, acrylates, quinones, reactive halides) with nucleophilic protein side-chains (lysine, serine, cysteine) *via* direct displacement

Table 3. Miscellaneous allergens *versus* non-allergens.

Case	Compound	Class	MR	PL	HBA	HBD
1	Griseofulvin	1	79.74	7.81	12	0
2	Penicillin G	1	86.26	4.98	10	2
3	Phthalic Anhydride	1	36.09	1.67	6	0
4	N-Methylol Chloroacetamide	1	26.90	1.14	5	2
5	Piperonal	1	38.44	2.21	6	0
6	m-Aminophenol	1	32.60	1.67	3	3
7	Benomyl	1	75.29	4.83	9	2
8	Captan	1	67.50	4.14	4	0
9	1,3-Dodecanesultone	1	65.76	5.93	6	0
10	Kitazin P	1	78.57	6.46	6	0
11	Picryl Chloride	1	50.38	2.15	6	0
12	2-Methoxy-4-propenylphenol	1	49.23	3.32	4	1
13	Dimethyl Isophthalate	0	47.43	3.26	8	0
14	Arachidonic Acid	0	98.62	9.61	4	1
15	3-Phenylpropionaldehyde	0	41.48	3.42	2	0
16	10-Undecen-1-ol	0	55.42	5.68	2	1
17	p-Cresyl Acetate	0	41.53	3.27	4	0
18	Benzoic Acid	0	32.29	1.90	4	1
19	Cholesterol	0	115.50	8.52	2	1
20	Glycerin	0	21.38	1.39	6	3
21	Leucine	0	35.45	2.70	5	3
22	Estradiol	0	78.69	4.57	4	2
23	Testosterone	0	85.37	6.73	4	1
24	Pregnane	0	90.84	6.32	0	0

Structural descriptors for modeling were of three types: continuous, ordinal of a leaving group, Schiff base formation or Michael addition. Prohapten activity is somewhat more complex as metabolic activity in the viable epidermis is a necessary step to generating a protein reactive hapten *in vivo*. In the equation to be described, IOH is a prohapten factor representing the action of a dermal dehydrogenase on a primary alcohol to generate a protein reactive aldehyde. The descriptor, IQUIN, represents the sensitivity of neutrally substituted phenols and anilines to undergo metabolic oxidase conversion to protein reactive quinones. Those phenols and anilines with oxidizable rings are coded, IQUIN = 1. We have observed that phenols and anilines with ortho-delocalizing groups such as nitro or carbomethoxy are exceptions to this rule and IQUIN = 0 for these compounds. The hapten descriptor, IPOS, is a composite descriptor representing several different types of electrophiles capable of reacting with the amino group of lysine or the mercapto group of cysteine. These include various conjugated systems, neutral sulfenylating agents such as captan and specially activated leaving groups. A full listing of these dichotomous descriptors has been published (*12*).

Early experiments with subsets of n = 24 indicated that linear discriminant analysis (LDA) would be a satisfactory technique. However, we have discovered an alternate discriminant method in two-class regression analysis which had been used successfully by Martin and coworkers(*14*). The method provides the same coefficients as LDA, but has many advantages in terms of descriptive plotting, identification of outliers and the development of an actual Y estimate or ACD score (values derived from the equation). The use of this technique to classify 36 allergens and 36 non-allergens has been described in detail (*12*). Only one statistical outlier was detected, though 5 compounds were incorrectly classified with respect to the classification boundary (0.50). The resulting model has the following form:

ACD response, CLASS = 1; Non-allergenic, CLASS = 0

	Coeff./Variable	T Value
CLASS =	0.00974 MR	2.10
	-0.153 PL	-3.21
	0.0468 HBA	2.52
	-0.154 HBD	-3.13
	-0.251 ICOOR	-3.09
	0.127 IX	3.40
	0.215 IOH	2.79
	0.564 IPOS	6.70
	0.465 IQUIN	6.01
	0.203 (constant)	

$n = 71 \quad r = 0.823 \quad s = 0.307 \quad F = 14.19$

Of the 9 descriptors, 2 each are continuous and ordinal. The remaining 5 represent statistical response to substructural descriptors. The continuous descriptor MR models polarizable volume or London forces and the coefficient is logically positive as size must be a factor in the T-cell recognition event. The best data indicate formaldehyde to be a variable allergen despite avid protein reactivity and this seems best explained by molecular size (*12*). The failure of LogP to achieve significance was surprising and has been verified by Cronin and Basketter using LDA with a completely different set of compounds (*15*). However, the lipophilic fraction of LogP (PL) proved to be a strong inhibitor of the ACD response, suggesting inappropriate partitioning into lipid domains. Both types of H-bonding were observed, indicating unspecified protein interactions. Of the dichotomous descriptors, only one was negative (ICOOR) and this can be interpreted as a degradation factor by aliphatic esterases in the epidermis. The others represent reactive halides (IX), metabolizable primary alcohols (IOH), a variety of electrophiles (IPOS) and phenols or anilines metabolizable to quinones (IQUIN). Other factors are surely present in this complex response, but these 9 represent the best statistical representation for this set of 71 compounds.

Plotting of Y estimates against CLASS (1.0 and 0.0) show the separation to be imperfect. There is an indeterminate range from 0.4-0.6 where classification must be considered unrealistic, as well as 2 misclassified allergens and 3 misclassified non-allergens. The overlap was anticipated as clinical testing on humans and animals is also imperfect due to a very broad range of skin types from sensitive to insensitive. Within the model, allergens are correctly classified at 94% and non-allergens at 92%. Extensive validation has been done with 80 chemicals that are far less well defined than those within the model and of much greater molecular variation. Allergens were correctly classified in 30/38 cases (79%) while nonallergens were correctly assigned in 37/42 cases (88%)(*12*). This success rate is comparable to that of clinical testing on a broad range of individuals.

Pesticide Skin Irritation and Dermatitis

The pesticide irritation ratings and categories of response to the ACD model are listed as follows. Under classification, IRR = immediate-type irritation, ACD = delayed-type allergic contact dermatitis, and NON = non-irritating by any mechanism.

Ratings	Classification
0.7-1.0	IRR(25)
0.7-1.0	ACD(9)
0.5	NON(7)
0.5	ACD(2)
0.0-0.3	NON(91)
0.0-0.3	ACD(14)

Nine of the 34 strongly irritating compounds are classified as ACD reactive. These consist of 4 neutral sulfenylating agents (captafol, captan, folpet and chlordantoin), 2 quinones (chloranil and dichlone) and 3 miscellaneous compounds which include the protein reactive acrolein. Only 2 of the 9 compounds are of questionable protein reactivity. One is etridiazole which contains a trichloromethyl group like DDT, but is much more hydrophilic and would not follow the same transport path as DDT. The other is vinclozolin which contains a cyclic carbamate of unknown reactivity. Both of these compounds were coded as protein reactive, but this is uncertain. The remaining 25 compounds did not classify as allergens and are considered to irritate by some form of immediate-type irritation (IRR).

The classification of the 9 compounds with ambiguous irritation comments is accepted as described by the model. The two compounds classified as ACD reactive are acrylonitrile and dichlofluanid. Both contain substructures that are clearly protein reactive and typical of other known ACD compounds (conjugated olefin and neutral sulfenylating agent).

The group of 105 pesticides rated non-irritating or only mildly irritating in The Agrochemicals Handbook (*1*) proved to be of exceptional interest. The model classifies 91/105 (87%) as nonallergens, a score comparable to earlier validation studies for this class (88%) and quite impressive as the model (n = 71) contains only five pesticides. Closer study of the 14 compounds classified as ACD shows only 4 with unambiguous protein reactivity. These are:

chlorothalonil	reactive ring halogens
halacrinate	conjugated olefin
pyridate	reactive ring halogen
pyridinitril	reactive ring halogens

Three other compounds coded for reactivity are ambiguous and may be false positives. Bupirimate contains a ring sulfate group, while triademefon has a potential triazole leaving group. Both are technically outside of our model building experience and were coded IPOS=1.0 with hesitation. Similarly, kadethrin has a conjugated olefin that would normally be coded for reactivity except for a bulky cyclopropyl group in the beta-position. As the expected protein reaction is conjugate addition of a thiol or amino group, the sensitivity of the Michael reaction to beta-substitution needs to be considered. Again, this situation is beyond our current experience and the coding for reactivity is questionable. Finally, bromofenoxim, an oxime related to bromoxynil and 5 organophosphates appear to have classified as ACD active for no other reason than possessing a HBA electron-pair count of 9-12. Clearly, none of these compounds will form 9-12 H-bonds simultaneously despite the electron-pair count. As the HBA coefficient is +0.047, these counts introduce a factor of 0.423-0.564 into the classification score. If one accepts these 9 compounds as ambiguous (4) or false positives (5), the classification performance of the model for non-irritating pesticides would increase to 91/96 (95%).

The four compounds identified as having ACD potential are highly probable and should be subjected to the guinea-pig maximization test (*16*) or the murine local lymph node assay (*6*) for confirmation. As contact dermatitis generally appears after sensitization (2nd challenge), ordinary irritation testing (single challenge) will not detect the immune response, nor are careful investigators likely to perceive it in laboratory studies. The problem can easily develope, however, through repetitive handling by the same manufacturing or farming personnel under less ideal conditions. Specific cases of pesticide allergy and other dermatoses in agricultural workers have been reviewed (*17*).

Summary

Of 148 pesticides with reported dermal effects, 139 were rated weak or strong. These were classified by a recently developed model for distinguishing non-allergens from compounds that induce allergic contact dermatitis. Six compounds were judged to be false positives and four to be of uncertain classification. This breaks down into 97 non-irritants (69.8%), 25 immediate-type irritants (18.0%), 13 contact dermatitis problems (9.4%) and 4 of unknown classification (2.9%). When we consider that the majority of these compounds were commercialized before extensive modern skin assays were routinely applied, the record of the pesticide industry looks very good. It is unlikely that new pesticides with severe ACD potential will reach the market under current toxicology standards. Some valuable new compounds will continue to be skin irritants, but this is far less serious than ACD reactivity and can be largely minimized by formulation and improved handling techniques.

Literature Cited

1. *The Agrochemicals Handbook*; Hartley, D. and Kidd, H., Eds. The Royal Society of Chemistry, Unwin Brothers Ltd.; Old Woking, Surrey, UK, **1983.**
2. von Blomberg, B.M.E.; Bruynzeel, D.P.; Scheper, R.J. in *Dermatotoxicology; Fourth Edition,* Marzulli, F.N.; Maibach, H.I., Eds.; Hemisphere Publishing Corp.: New York, **1991**, Chapter 12.
3. Dupuis, G.; Benezra, C. *Allergic Contact Dermatitis to Simple Chemicals;* Marcel Dekker, Inc.: New York and Basel, **1982**, Chapters 5 and 6.
4. Thornfeldt, C.R., dermatologist, private communication: Bircher, A.J.; Howald, H.; Rufli, Th., *Contact Dermatitis* **1991**, *25*, 230-236.
5. Reference 3, p. 82.
6. Kimber, I.; Weisenberger, C., *Arch. Toxicol.* **1989**, *63*, 274-282.
7. Hostynek, J.J.; Lauerma, A.I.; Magee, P.S.; Bloom, E.; Maibach, H.I., *Acta Dermato-Venerealogica*, accepted for publication.
8. Reese, K.M., *Chem. Engr. News* **1994**, *April 25*, 74.
9. Ziegler, V.; Richter, C.; Ziegler, B.; Uwe-Frithjof, in *Seminars in Dermatology*, Maibach, H.I., Ed.; W.B. Saunders Company, Philadelphia, **1989**, Vol. 8, No.2, 80-82.
10. *Catalog of Givaudan Aroma Chemicals,* **1984** ,Givaudan-Roure.
11. *CTFA Cosmetic Ingredient Dictionary; Fourth edition,* Nikitakis, J.M.; McEwen, Jr., G.N. and Wenninger, J.A., Eds.; The Cosmetic, Toiletry, and Fragrance Association: Washington, DC., **1991.**
12. Magee, P.S.; Hostynek, J.J.; Maibach, H.I., *Quant. Struct.-Act. Relat.* **1994**, *13*, 22-33.
13. Magee, P.S. in *The Science of the Total Environment;* Hermens, J.L.M.; Opperhuizen, A., Eds.; Elsevier Science Publishers, B.V.: **1991**, Amsterdam, pp. 155-178.
14. Martin, Y.C.; Holland, J.B.; Jarboe, C.H.; Plotnikoff, N., *J. Med. Chem.* **1974**, *17*, 409.
15. Cronin, M.T.D.; Basketter, D.A. in *Trends in QSAR and Molecular Modeling 92*; Wermuth, C.D., Ed.; ESCOM: Leiden, **1993**; pp. 297-298.
16. Wahlberg, J.E.; Boman, A. in *Current Problems in Dermatology*, Vol. 14; Andersen, K.E.; Maibach, H.I., Eds.; Karger: Basel, **1985**, pp. 59-106.
17. Abrams, K.; Hogan, D.J.; Maibach, H.I. in *Occupational Medicine: Health Hazards of Farming,* Cordes, D.H. and Rea, D.F., Eds.; Hanley & Beltus, Inc.: **1991**, Philadelphia, Vol. 6, pp. 463-492.

RECEIVED July 28, 1995

Chapter 10

QSARs in Environmental Toxicology and Chemistry

Recent Developments

Joop L. M. Hermens and Henk J. M. Verhaar

Research Institute of Toxicology, Utrecht University, P.O. Box 80176, 3508 TD Utrecht, Netherlands

If used within their limitations, QSARs can be extremely useful for evaluating the environmental risks of the large number of existing chemicals. This paper will focus on the practical application of QSARs as well as on some new developments in the field of environmental QSAR, such as: the use of multivariate techniques and the application of quantum-chemical parameters. PLS models combined with quantum-chemical parameters form a good combination, not only for the development of QSARs, but also for the interpretation of models in mechanistic terms.

The first QSARs originate from studies to the biological activity of agrochemicals (*1*). During the last decade, the number of QSAR studies in the field of environmental chemistry and ecotoxicology has been growing steadily (*2-8*). The reason for this growing interest in using QSAR for environmental parameters is certainly related to the large number of chemicals which has to be regulated. For example, in 1993 the European Regulation on the Evaluation and the Control of the Environmental Risks of Existing Substances came into force (*9*). Within this regulation an extremely large number of chemicals has to be regulated; the European Inventory of Existing Chemical Substances (EINECS), for example, lists more than 100.000 substances. Not only because of time constraints, but also because of limited financial resources, it is impossible to test all these existing chemicals for their possible adverse effects on the environment. Structure-activity relationships can form an instrument to fill these data gaps. It is not surprising that at this moment several national and international activities are being undertaken to assess the potential use of QSARs in the evaluation of new as well as existing chemicals (*10-13*). To this purpose, compilations of QSARs for environmental parameters are extremely useful and they can be found in several publications, reports or computer programs from different organizations or research groups in the US and Europe (*14-16*).

0097–6156/95/0606–0130$12.00/0

When QSAR is used in a regulatory context, it becomes important to define the applicability and *limitations of a QSAR* more carefully because each predictive model, including QSAR, has its limitations. This issue is addressed more extensively in the first part of this chapter. Techniques to establish QSARs are continuously under development. In general, new developments originate from QSAR research in medicinal chemistry and agrochemistry. Two developments in environmental QSAR, namely *the application of multivariate techniques and quantum-chemical properties*, are addressed in the second part of this manuscript.

Limitations of QSARs

Numerous QSARs have been published for all types of different environmental parameters such as chemical degradation, biodegradation, bioconcentration and ecotoxicity. In most cases, however, the limitations of the presented models are not carefully, if at all, presented in these publications. This is not due to the unwillingness of the authors. In many cases, the models are published not directly with the intention to have them applied in practice but because they are interesting from a scientific point of view; they may e.g. supply some mechanistic information on the process.

In defining the *limitations of a certain QSAR*, two different aspects can be distinghuished: (i) the uncertainty in the model itself, and (ii) the chemical domain of the model; in other words: for which type of chemicals is the model applicable ? The first aspect is in principle the easiest because it can be presented by a simple number such as a standard error of estimate or even better by a r^2 or a cross validated Q^2. Defining the chemical domain of a model, however, is much more complex. In principle, two options can be followed in defining the domain: (i) a multivariate profile of the chemical descriptor space, or (ii) an expert judgment based set of structural rules or a combination of (i) and (ii). The need for defining the structural requirements of a certain model has been recognized before and was also the subject of an EPA workshop held in 1988 (*17*).

A relatively vague description of the domain in expert rules is usually not sufficient because very small changes in chemical structure may result in a completely different behaviour. A definition of a domain such as by stating that "this acute fish toxicity model is valid for halogenated nitroaromatics or is valid for aliphatic chlorohydrocarbons" is too broad and needs more specific information. Table I gives two examples in order to illustrate this. Chemicals such as 3,5-dichloronitrobenzene (A) and 1,2-dichloropropane (C) have a relatively low acute fish toxicity, while 1-chloro-2,4-dinitrobenzene (B) and 2,3-dichloropropene (D) are much more toxic. Although the chemical structures of chemicals A and B, and chemicals C and D, are very similar, chemicals B and D are alkylating agents which explains their relatively high toxicity. It is also obvious from these simple examples that information on mechanisms of the process will strengthen the possibility to define the limitations of a certain QSAR (*18,19*).

An Example for Fish LC_{50} QSARs. Fish LC_{50}s of large sets of organic micropollutants are almost perfectly related to their octanol-water partition coefficient (K_{ow}) (*6,20-23*). These QSARs are only valid for relatively unreactive chemicals with a non-specific mode of action, and they represent a so-called base-line toxicity.

Table I. 14-day LC_{50} of Four Chemicals (from Hermens (*6*))

chemical	log K_{ow}	LC_{50} (mol/L)
chlorinated nitroaromatics		
A. 3,5-dichloronitrobenzene	3.20	$2.3\ 10^{-5}$
B. 1-chloro-2,4-dinitrobenzene	2.20	$6.5\ 10^{-7}$
aliphatic chlorohydrocarbons		
C. 1,2-dichloropropane	2.16	$1.0\ 10^{-3}$
D. 2,3-dichloropropene	1.99	$1.0\ 10^{-5}$

Therefore, this class of chemicals is also called "class I" compounds. Very similar relationships have been established for other endpoints and species. Examples are presented in Figure 1a for LC_{50} to guppy (*20*) and in Figure 1b for No-observed Effect Concentrations (NOECs) for 19 different aquatic organisms (*23*). As shown in Figure 2b, differences in sensitivity of aquatic organisms for class I chemicals are relatively small, and this is to be expected for unspecific toxicity. This example is of course not new and the concepts go back to the early work of Hansch and Glave (*24*), Hansch and Fujita (*25*) or even to the end of the last century (*26*).

The point we would like to make here is the definition of the limitations of these models. It has been known for a long time, that more polar chemicals (*27,28*) (class II), reactive chemicals (*6,22,29,30*) (class III) as well as chemicals with a specific mode of action (*6,40*) (class IV) cause mortality at much lower concentrations than their corresponding base-line toxicity and examples are presented in Figure 2. The important question is then: how can we indicate for which chemicals the QSAR for unspecific toxicity can be applied ? In other words: which type of chemicals belong to class I and the other three classes. A strict definition of chemical structures within each class is the most appropriate way and a few examples of the chemical domain of class I chemicals are given in Figure 2. The indication of the chemical domain in Figure 2 is rather broad, but a very detailed and extensive compilation of structural rules for these classes has been given (*31*). When these rules are followed strictly, these QSARs can be applied to estimate aquatic effect concentrations for those chemicals which fall within the domain. Aquatic toxicity QSARs for class I chemicals were recently applied to the High Production Volume Chemicals (HPVCs) of the EC Inventory (*32*). It appeared that around 25 % of the discrete HPVSs on this inventory could be classified as narcosis type chemicals. Application of the models to the whole EINECS will result in predictions of the aquatic toxicity for thousands of chemicals showing the tremendous value of QSAR for filling data gaps.

As indicated above, it is necessary to evaluate the applicability of estimation models, including their limitations. The basic prerequisite for such an evaluation is the preparation of overviews of existing models. The preparation of such overviews is also an objective and task in an international project entitled "Quantitative Structure-Activity Relationships (QSARs) for Predicting Fate and Effects of Chemicals in the

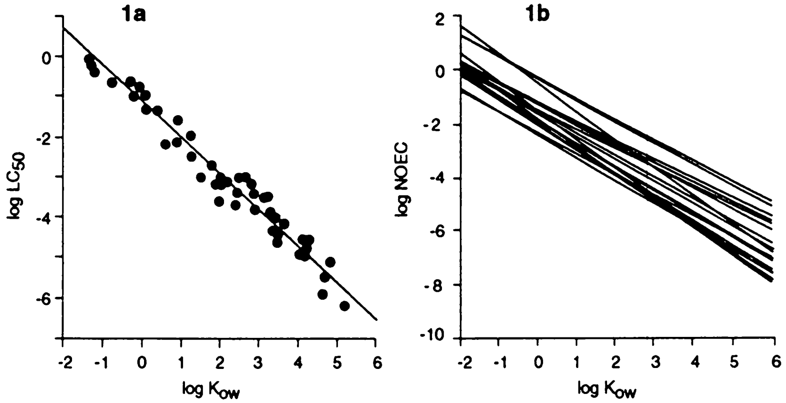


Figure 1. Relationship between LC_{50} to guppy and K_{ow} for chemicals with unspecific mode of action. (a) LC_{50} (mol/L) to guppy (data from Könemann (*20*)), (b) No-observed effect concentrations (mol/L) for 19 species (data from van Leeuwen et al. (*23*)).
Data presented in figure 1a are from ref. 20; figure 1b is reproduced with permission from ref. 23.

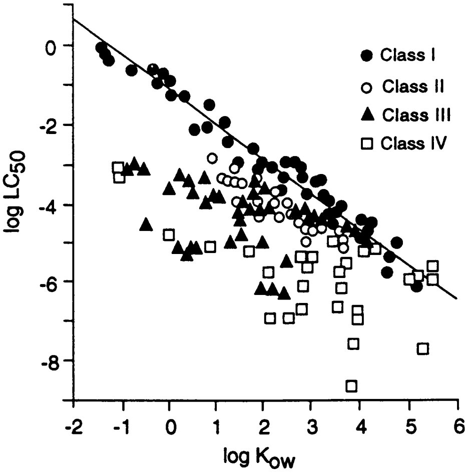
Environment". This project, in which six European institutes are participating, is being carried out in the framework of the Environmental Technologies RTD Programme (DG XII/D-1) of the Commission of the European Union.

Multivariate Approaches and Quantum-Chemical Parameters

The Importance of Chemical Reactivity. The octanol-water partition coefficient (K_{ow}) is one of the most frequently used parameter in QSARs. This parameter reflects hydrophobicity, and in the absence of specific type of interactions, several processes are related to K_{ow}. It should come as no surprise that QSARs for bioconcentration factors, soil sorption coefficients and ecotoxicity of neutral organic compounds usually contain only K_{ow} as the most relevant parameter.

If specific types of interactions occur, whether they be electronic or steric interactions, K_{ow} will not be sufficient and other parameters are needed. Specific interactions are less frequent in processes such as soil sorption (although they do of course occur), but in chemical degradation, biodegradation and biotransformation such specific interactions are more common. QSARs in medicinal- and agrochemistry are usually established for specific interactions because such chemicals are designed for their specificity. In QSARs for drugs, as well as for agrochemicals, the use of stereo-electronic parameters, or techniques to model specific interactions has therefore a much longer tradition, with the application of quantum-chemical parameters or approaches such as CoMFA as examples (*33*).

While the first QSAR studies for environmental endpoints were mainly focused on relatively non-polar or neutral organic compounds, environmental QSAR research is now dealing with more diverse chemical structures. It is because of this that the



Class I: aliphatic and aromatic (halogenated) hydrocarbons; Class II: aromatic amines, nitroaromatics and phenols; Class III: epoxides, aldehydes, aziridines, quinones (in general: all alkylating agents); Class IV: organophosphates, chlorinated pesticides etc. (in general: all specifically acting chemicals). A very detailed description of the chemical domain of these classes is given by Verhaar et al. (*31*).

Figure 2. LC_{50} (mol/L) to guppy of chemicals from different classes and some examples of the chemical domain of these models (from Verhaar et al. (*31*)).

number of QSAR for environmental processes which include other descriptors besides K_{ow} is increasing. In particular, quantum-chemical descriptors are nowadays applied frequently. One of the advantages of quantum-chemical parameters is of course that they can be calculated directly from structure.

In principle, different approaches can be followed in using quantum-chemical techniques. It is obvious that the most appropriate way of using quantum-chemical parameters is by selection of a certain parameter in advance based on knowledge of the mechanisms of the process. Calculation of the activation energy of a reaction, needed to form a transition-state, is an example of a mechanistic approach because this is directly related to the reaction rate constant. However, the calculation of the geometry and energy parameters, including activation energy, of the transition-state is far more complex than the calculation of ground-state properties.

In many cases, the exact mechanisms involved are unknown and in those cases multivariate techniques can be of use, in particular because semi-empirical MO calculations can yield a lot of different parameters which are often highly collinear. In particular, PCA and PLS analyses constitute appropriate tools to analyse these large number of (possible collinear) parameters. More information on these techniques can be found in several papers by Wold and coworkers (*34,35*) and examples in the field of environmental sciences are given by Tosato et al. (*36*) and Eriksson et al. (*37*). Some examples of QSARs for environmental endpoints are summarized in Table II.

Table II. Some Examples of QSAR Studies for Environmental Endpoints Based on Quantum-Chemical Parameters

process	quantum-chemical parameters (technique)[a]	reference
fish toxicity of organophosphates	several ground-state parameters (MLR)	Schüürmann (*41*)
reaction of epoxides with a nucleophile	superdelocalizability (MLR)	Purdy (*42*)
fish toxicity of polar chemicals	superdelocalizability index (MLR)	Veith et al. (*43*)
reaction of epoxides with a nucleophile	several ground state parameters (PLS)	Eriksson et al. (*44*)
reaction of hydrocarbons with OH radical	ground-state parameters (PLS)	Eriksson et al. (*45*)
fish toxicity, bioconcentration and biotransformation of organophosphates	several ground-state parameters (PLS)	Verhaar et al. (*46*)

a) MLR: Multiple Linear Regression, PLS: Partial Least Squares projection to latent structures analysis.

The combination of multivariate techniques and quantum-chemical parameters was successfully applied to reactivity data of epoxides, to reaction rates of chlorinated aliphatic hydrocarbons with hydroxyl radical, and to a series of biological endpoints for organophosphates. This last example is discussed here in more detail.

An Example for Organophosphates. De Bruijn et al. (*38-40*) determined several different biological endpoints for a series of organophosphates; from LC_{50}s, bioconcentration and biotransformation in fish to inhibition of the target enzyme: acetylcholinesterase (AChE). As shown in Figure 3 all these processes affect the actual toxicity of these pesticides.

Nine different quantum-chemical parameters were used, together with log K_{ow}, by Verhaar et al. (*46*) in a PCA/PLS analysis for developing models for all endpoints. The models for two endpoints (bioconcentration: BCF) and inhibition of AChE), are discussed here in more detail.

Model for Bioconcentration The previous work of de Bruijn et al. (*38*) has shown that bioconcentration factors (BCFs) of most of the tested organophosphates agree with the behaviour of neutral organic chemicals: i.e. BCF (if expressed on a lipid weight basis) is equal to K_{ow} (see Figure 4). However, it is also obvious from Figure 4 that five of the organophosphates have substantially lower BCF values. These five chemicals are chlorothion, dicapthon, iodofenphos, ronnel and bromophos. The lower BCF values for these five chemicals are likely related to their (relatively) rapid biotransformation via dealkylation; in vitro biotransformation studies in fish support this hypothesis (*39*). The result of the PLS model for BCF is given in Table III in the form of pseudo-regression coefficients.

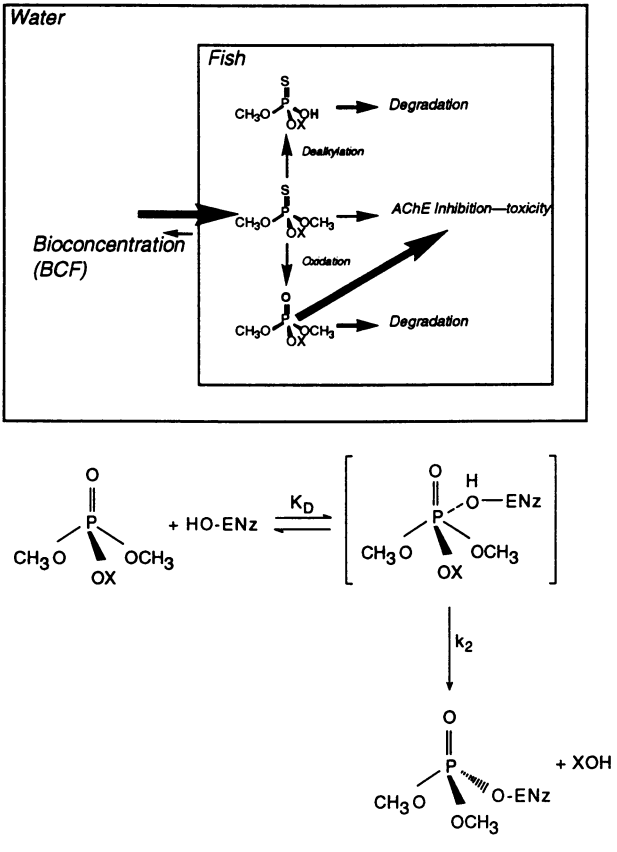

Water
Fish
Degradation
Dealkylation
AChE Inhibition—toxicity
Bioconcentration
(BCF)
Oxidation
Degradation
+ HO-ENz
K_D
k_2
+ XOH

Figure 3. Overview of processes which determine the toxicity of organophosphates and the two steps in the inhibition of the enzyme acetylcholin-esterase (AChE).

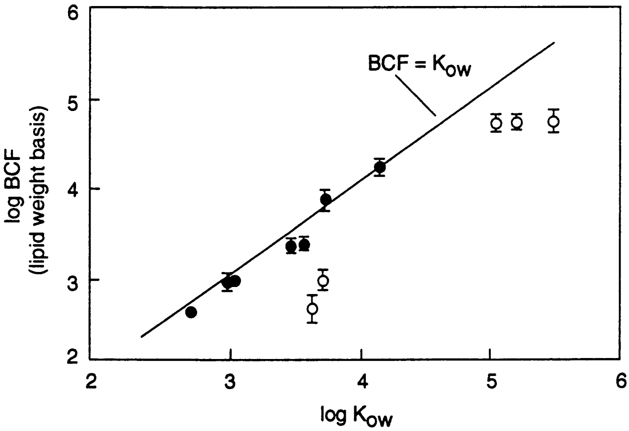


Figure 4. Bioconcentration factors (BCFs) in guppy (*Poecilia reticulata*) of 12 organophosphates versus the octanol-water partition coefficient (K_{ow}). Open circles are those organophosphates which are biotransformed relatively fast. Figure adapted from ref. 38.

Table III. PLS Models (in the Form of Regresssion Coefficients) for the Bioconcentration in Fish and the two Steps in the Inhibition of AChE by a Series of Organophosphates (Verhaar et al. (*46*))

Chemical parameter[a)]	Pseudo-regression coefficient[b)]		
	BCF	K_D	k_2
E_{homo}	0.046	-0.016	0.126
E_{lumo}	**0.198**	0.095	**-0.811**
EN	-0.085	-0.011	0.103
n	0.023	0.063	**-0.515**
$D^N(P)$	0.003	0.063	-0.428
$D^E(S)$	0.017	-0.040	0.296
q(P)	0.020	-0.127	0.117
p(PO)	-0.009	**-0.260**	**-0.512**
delta q(PO)	0.022	-0.193	0.240
log K_{ow}	**0.898**	**-1.024**	-0.438

a) E_{homo}: energy of the highest occupied molecular orbital, E_{lumo}: energy of the lowest unoccupied molecular orbital, EN: electronegativity, *n*: absolute hardness, $D^N(P)$: nucleophilic delocalizability on phosphorus atom, $D^E(S)$: electrophilic delocalizability on sulfur atom, q(P): charge on phosphorus atom, p(PO): bond order of the phosphorus-oxygen bond, delta q(PO): charge difference between the phosphorus and oxygen atom, K_{ow}: octanol-water partition coefficient.

b) The PLS model can be rewritten in the form of a multiple linear regression equation. The coefficients are scaled and mean-centered and their value supply information on the influence of the original X-variable. The two or three most important parameters (highest coefficients) are highlighted.

For bioconcentration, K_{ow} is the most predominant parameter and the next important one is E_{lumo}. The influence of K_{ow} is not surprising, but an influence of E_{lumo} on BCF is not common. However, for these chemicals, the role of this second parameter may very likely be related to its correspondence with the rate of biotransformation. E_{lumo} is probably the parameter which models the de-alkylation rates of the organophosphates by glutathione-S-transferase; de-alkylation is an attack of a nucleophile (glutathione).

In conclusion: both the simple linear regression approach (Figure 4) and the multivariate model (Table III) lead to similar conclusions and they also support each other.

Model for AChE Inhibition In the PLS model (see Table III) for the first step in the inhibition (the formation of the enzyme-substrate complex) K_{ow} as well as the bond order of the P-O bond are predominant, while for the second step, the E_{lumo} dominates. The influence of K_{ow} in the first step of inhibition suggest that hydrophobic interactions are important in association between the organophosphate and the enzyme. So, besides the fact that PLS generally results in more stable models, PLS can also yield information which is of relevance for the mechanistic interpretation of a certain process.

Both examples show that PLS models and quantum-chemical parameters form a good combination, not only for the development of QSARs but, also for the interpretation of models in mechanistic terms.

QSAR is not only a predictive technique, but it is also a way of looking at experimental data and, as pointed out by Fujita (*47*), QSARs may result in new leads. Moreover, QSAR may give hints with respect to mechanistic interpretation in particular, if, as proposed by Hansch (*48*) QSARs for different endpoints are compared. This is also the case for the, sometimes, relatively simple endpoints in environmental QSAR.

Acknowledgements

A large part of the work described here is being carried out within the framework of the project "Quantitative Structure-activity Relationships for Predicting Fate and Effects of Chemicals in the Environment". This project is supported financially by the Environmental Technologies RTD Programme (DG XII/D-1) of the Commission of the European Union under contract number EV5V-CT92-0211. This financial support from the European Union is gratefully acknowledged.

References

1. Hansch, C. In *Structure-Activity Relationships, vol 1;* Cavallito, C.J., Ed.; Pergamon Press: Oxford, 1973; pp. 75-165.
2. Turner, L.; Choplin, F.; Dugard, P.; Hermens, J.; Jaeckh, R.; Marsmann, M.; Roberts, D. *Toxic. in Vitro* **1987**, *1*, 143-171.
3. *QSAR in Environmental Toxicology-II;* Kaiser, K.L.E., Ed.; Reidel: Dordrecht, 1987.

4. Nirmalakhandan, N.; Speece, R.E. *Environ. Sci. Technol.* **1988**, *22*, 606-615.
5. *Practical Applications of Quantitative structure-activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., Eds.; Kluwer: Dordrecht, 1990.
6. Hermens, J. In *Handbook of Environmental Chemistry, vol. 2E.*; Hutzinger, O., Ed.; Springer: Berlin, 1989; pp. 111-162.
7. *QSAR in Environmental Toxicology IV*; Hermens, J.; Opperhuizen, A., Eds.; Elsevier: Amsterdam, 1991.
8. Calamari, D.; Vighi, M. *Rev. in Environ. Toxicol.* **1990**, *4*, 1-122.
9. *Council Regulation (EEC) on the Evaluation and the Control of the Environmental Risks of Existing Substances.* No. 783/93.
10. *Report of the OECD Workshop on the Application of QSARs in Aquatic Effects Assessment*; OECD Environment Monograph No. 58, OECD: Paris, 1992.
11. *A collaborative US-EPA-EU project on fate and effects predictions for High Production Volume Chemicals*; European Chemicals Bureau, Joint Research Centre: Ispra, Italy, 1994.
12. Hermens J., Balaz S., Damborsky J., Karcher W., Müller M., Peijnenburg W., Sabljic A. and Sjöström M. Submitted for publication.
13. *Guidance Document on the Use of QSAR in Risk Assessment.* European Union: Brussels, 1994.
14. Clements R.G.; Nabholz, J.V. *ECOSAR: A computer program for estimating the ecotoxicity of industrial chemicals based on structure-activity relationships*; US-EPA Environmental Effects Branch. US-EPA. report number 748-R-93-002, 1994.
15. *Handbuch zum SAR-Programm Version 3.0*; Fraunhofer-Institut für Umweltchemie und Ökotoxikologie: Schmallenberg, Germany, 1992.
16. Russom, C.L.; Anderson, E.B.; Greenwood, B.E.; Pilli, A. *Sci. Total Environ.* **1991**, *109/110*, 667-670.
17. Bradbury, S.P.; Lipnick. R.L. *Environ. Health Perspect.* **1990**, *87*, 181-182.
18. McKim, J.M.; Bradbury, S.P.; Niemi, G.J. *Environ. Health Perspect.* **1987**, *71*, 171-186.
19. Bradbury, S.P.; Henry, T.R.; Carlson, R.W. In *Practical Applications of Quantitative structure-activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., Eds.; Kluwer: Dordrecht, 1989; pp. 295-315.
20. Könemann, H. *Toxicology* **1981**, *19*, 209-221.
21. Veith G.D.; Call, D.J.; Brooke L.T. *Can. J. Fish. Aquat. Sci.* **1983**, *40*, 743-748.
22. Lipnick, R.L.; Watson, K.R.; Strausz, A.K. *Xenobiotica* **1987**, *17*, 1011-1025.
23. Van Leeuwen, C.J.; van der Zandt, P.T.J.; Aldenberg, T.; Verhaar, H.J.M.; Hermens, J.L.M. *Environ. Toxicol. Chem.* **1992**, *11*, 267-282.
24. Hansch, C.; Glave, W.R. *Mol. Pharmacol.* **1971**, *7*, 337-354.
25. Hansch, C; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.
26. Lipnick, R.L. *Trends in Pharmacol. Sci.* **1986**, *7*, 161-164.
27. Veith, G.D.; Broderius, S.J. In *QSAR in Environmental Toxicology-II*; Kaiser, K.L.E., Ed.; Reidel: Dordrecht, 1987; pp. 385-391.
28. T.W. Schultz; Holcombe, G.W.; Phipps, G.L. *Ecotoxicol. Environ. Safety* **1986**, *12*, 146-153.

29. Hermens, J. *Environ. Health Pespect.* **1990**, *87*, 219-225.
30. Veith, G.D.; Lipnick, R.L.; Russom, C.L. Xenobiotica **1989**, *19*, 555-565.
31. Verhaar, H.J.M.; van Leeuwen, C.J.; Hermens, J.L.M. *Chemosphere* **1992**, *25*, 471-491.
32. Verhaar, H.J.M.; Leeuwen, C.J.; Bol, J.; Hermens, J.L.M. *SAR and QSAR in Environ. Res.* **1994**, *2*, 39-58.
33. Cramer, R.D., III; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
34. Wold, S.; Dunn, W.J. *Chem. Inf. Comput. Sci.* **1983**, *23*, 6-13.
35. Wold, S.; Dunn, W.J.; Hellberg, S. *Environ. Health Perpect.* **1985**, *61*, 257-268.
36. Tosato, M.L.; Chiorboli, C.; Eriksson, L.; Jonsson, J.; Marchini, S.; Passerini, L.; Pino, A.; Vigano, L. *J. Am. Coll. Toxicol.* **1990**, *9*, 629-638.
37. Eriksson, L.; Jonsson, J.; Hellberg, S.; Lindgren, F.; Skagerberg, B.; Sjostrom, M.; Wold, S.; Berglind, R. *Environ. Toxicol. Chem.* **1990**, *9*, 1339-1351.
38. De Bruijn, J.; Hermens, J. *Environ. Toxicol. Chem.* **1991**, *10*, 791-804.
39. De Bruijn, J.; Seinen, W.; Hermens, J. *Environ. Toxicol. Chem.* **1993**, *12*, 1041-1050.
40. De Bruijn, J.; Hermens, J. *Aquat. Toxicol.* **1993**, *24*, 257-274.
41. Schüürmann, G. *Environ. Toxicol. Chem.* **1990**, *9*, 417-428.
42. Purdy, R. *Sci. Total Environ.* **1991**, *109/110*, 553-556.
43. Veith, G.D.; Mekenyan, O.G. *Quant. Struct.-Act. Relat.* **1993**, *12*, 349-356.
44. Eriksson, L.; Verhaar, H.; Hermens, J. *Environ. Toxicol. Chem.* **1994**, *13*, 683-691.
45. Eriksson, L.; Rännar, S.; Sjöström, M.; Hermens, J.L.M. *Environmetrics* **1994**, *5*, 197-208.
46. Verhaar, H.J.M.; Eriksson, L.; Sjöström, M.; Schüürmann, G.; Hermens, J.L.M. *Quant. Struct.-Act. Relat.* **1994**, *13*, 133-143.
47. Fujita, T. In *QSAR in Design of Bioactive Compounds;* Kuchar, M., Ed.; J.R. Prous Science, 1992; 3-22.
48. Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147-153.

RECEIVED April 26, 1995

Chapter 11

Noncongeneric Structure–Toxicity Correlation Using Fuzzy Adaptive Least-Squares

I. Moriguchi[1], Q. Liu[1], H. Hirano[2], and S. Hirono[1]

[1]School of Pharmaceutical Sciences, Kitasato University, Shirokane, Minato-ku, Tokyo 108, Japan
[2]Zeria Pharmaceutical Company, Ltd., Oshikiri, Konan-machi, Saitama 360–01, Japan

Quantitative relationships between chemical structure and toxicity of miscellaneous organic chemicals were studied to generate predictive models for screening compounds for possible environmental risks and pollutant hazards. The toxicity data included the aquatic toxicity of 394 compounds, rodent carcinogenicity and mutagenicity of 246 compounds, and biodegradability of 1259 compounds. The non-congeneric structure-toxicity relationships were analyzed using fuzzy adaptive least-squares (FALS), a pattern recognition method recently developed in our laboratory for structure-activity rating correlation. A novel feature of FALS is that the degree to which each sample belongs to its activity class is given by a fuzzy membership function. The generated QSAR models were statistically reliable in both recognition and leave-one-out prediction despite the diversity and complexity of the structures of the compounds investigated.

Biological activity and toxicity data are often described in the form of activity ratings. Such data include *in vivo* data of low accuracy, rough data from a single dose response, and a large amount of data assembled from more than one source. For the correlation of molecular structure with activity ratings, we developed the adaptive least squares (ALS) method (*1–3*), which has been successfully used for over a decade. Activity ratings comprise not only statistical vagueness such as measurement inaccuracy but also intrinsic vagueness such as individual differences in a living body and subjective criteria for classification. Such indefiniteness can be treated by the concepts of fuzzy variance. According to the technique of the fuzzy inference (*4*), we have introduced a membership function to ALS to develop the fuzzy ALS (FALS) (*5–9*). A novel feature of FALS is that the degree to which each sample belongs to its activity class is given by the fuzzy membership function.

In recent years, we have studied the quantitative relationships between chemical structure and toxicity of non-congeneric organic chemicals using FALS to generate predictive models for screening compounds for possible health and environmental hazards. Since sets of toxicity data for a non-congeneric QSAR study are usually

0097–6156/95/0606–0141$12.00/0

compiled from more than one source, the use of FALS is considered suitable. In the present study, the toxicity data included the aquatic toxicity of 394 compounds mainly cited from publications by the US National Technical Information Service, rodent carcinogenicity and mutagenicity of 246 compounds tested by the US National Toxicology Program, and biodegradability of 1259 compounds collected from the JETOC/KASHIN database. The generated QSAR models were statistically reliable in both recognition and leave-one-out prediction despite the diversity and complexity of the structures of the compounds investigated.

FALS Method

FALS is a nonparametric pattern classifier and formulates QSAR in a single equation (equation 1), irrespective of the number of activity rating classes using an error-correcting feedback adaptation.

$$Z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_px_p \qquad (1)$$

In this equation, x_k = kth descriptor (k = 1,2,...,p) for structures, w_k (k = 0,1,2,...,p) = weight coefficient, and Z = discriminant score. In the m-class discrimination of n samples, the starting scores, a_j (j = 1,2,...,m), for the members of class j are assumed as equation 2, and the class boundaries, b_j (j = 1,2,...,m−1), based on the Z value are fixed as the midpoints between the starting scores a_j and a_{j+1}.

$$a_j = 4\left(\sum_{g=1}^{j-1} n_g + n_j/2\right) / n-2 \qquad (2)$$

In equation 2, n_g and n_j are the number of samples in classes g and j, respectively. Classes are usually numbered in ascending order of potency. A bell-shaped membership function, $M_j(Z)$ for class j (equation 3), is assumed to give the membership grade of classes for the compounds.

$$M_j(Z) = \begin{cases} 1/[1+\{(Z-b_{j-1})/Fl-1\}^4] & Z \le b_{j-1}+Fl \\ 1 & b_{j-1}+Fl < Z \le b_j - Fl \\ 1/[1+\{(b_j-Z)/Fl-1\}^4] & b_j - Fl < Z \end{cases} \qquad (3)$$

Additionally, $M_j(Z)$ is taken to be 1 when $Z \le b_1 - Fl$ for class 1 and when $Z > b_{m-1} + Fl$ for class m. Fl is the parameter for fuzziness of the boundary between classes, usually being taken to be 0.1. The value of the membership function, which is called the membership grade, MG, ranges from 0 to 1 and is taken to be 0.5 at the class boundaries.

In FALS 91 (*6–9*), the weight coefficients in equation 1 are iteratively estimated through a two-step adaptive least squares calculation: Step 1 ($t \le 20$) for a rough but wide search to avoid falling into a local optimum and Step 2 ($t > 20$) for weighted least squares optimization, as shown in Figure 1. The value t is the iteration time. In Figure 1, RM is the product of R_S (Spearman's rank correlation coefficient) and MMG (mean membership grade). In both steps, forcing factors, S_i (i = 1,2,...,n), are assumed in place of Z in equation 1 for n compounds, and w_k are calculated using a least squares method. The starting scores (equation 2) are used for S_i at iteration 1. At iteration 2 and thereafter, the forcing factor $S_i^{(t+1)}$ ($t \geq 1$) is adapted using the correction term $C_i^{(t)}$ as

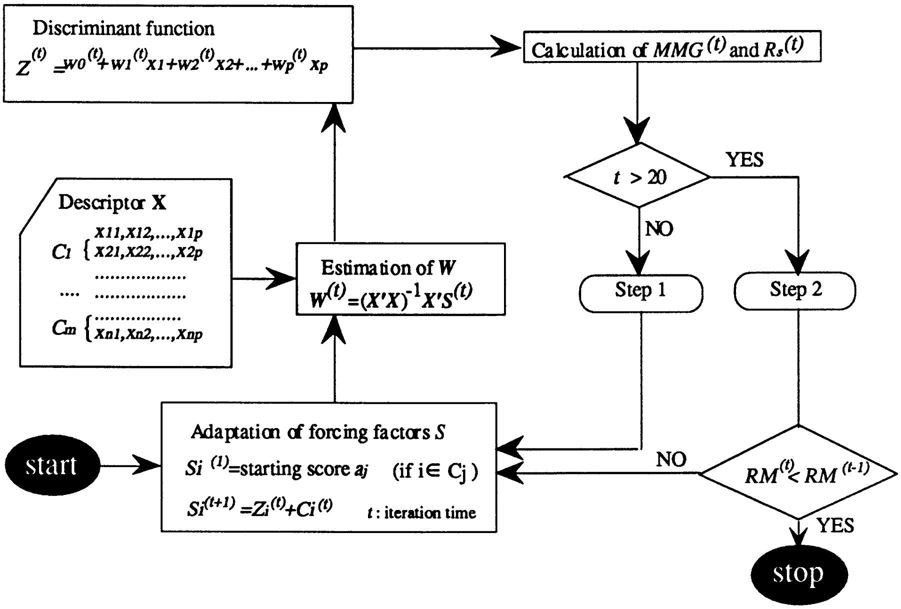

Discriminant function
Z(t) = w0(t) + w1(t)x1 + w2(t)x2 + ... + wp(t)xp
Calculation of MMG(t) and Rs(t)
t > 20
YES
NO
Step 1
Step 2
Descriptor X
x11,x12,...,x1p
C1 {x21,x22,...,x2p
Cn {xn1,xn2,...,xnp
Estimation of W
W(t) = (X'X)-1X'S(t)
Adaptation of forcing factors S
Si(1) = starting score aj (if i ∈ Cj)
Si(t+1) = Zi(t) + Ci(t)
t : iteration time
start
NO
RM(t) < RM(t-1)
YES
stop

Figure 1. Process of FALS Calculation
Reproduced with permission from ref. 9.
Copyright 1994 Pharmaceutical Society of Japan.

$$S_i^{(t+1)} = Z_i^{(t)} + C_i^{(t)} \tag{4}$$

In Step 1, the ALS correction term (equation 5) (*3, 6–8*) is used for $C_i^{(t)}$.

$$C_i^{(t)} = \begin{cases} 0.1/(\delta_i^{(t)}+\alpha)^2 + 0.1 & Z_i^{(t)} < b_{j-1} \\ 0 & b_{j-1} \leq Z_i^{(t)} \leq b_j \\ -0.1/(\delta_i^{(t)}+\alpha)^2 - 0.1 & b_j < Z_i^{(t)} \end{cases} \tag{5}$$

Here, $\delta_i^{(t)}$ (> 0) is the distance between $Z_i^{(t)}$ and the nearer boundary, b_{j-1} or b_j, for compound i which is a member of class j, and α is the constant usually taken to be 0.45. A set of weight coefficients giving the best discrimination within 20 iterations is selected as the initial coefficients in step 2.

In Step 2, the following correction term (equation 6) is used for $C_i^{(t)}$,

$$C_i^{(t)} = \begin{cases} \beta\sqrt{(1-MG_i^{(t)})\,Fl} & Z_i^{(t)} \leq a_j \\ -\beta\sqrt{(1-MG_i^{(t)})\,Fl} & Z_i^{(t)} > a_j \end{cases} \tag{6}$$

where β is the constant usually taken to be 0.5. The adaptive least squares calculation is carried out so as to minimize $\Sigma(S_i - Z_i)^2$, or ΣC_i^2 from equations 1 and 4. Therefore, as is obvious from equation 6, we can expect to obtain a discriminant function giving maximum ΣMG_i over the set of n compounds in this step. As the criterion for the best discrimination in Step 2, the product of R_S and MMG is used; R_S supplements the information about overall accuracy of discrimination. The Step 2 iteration is performed until the criterion is no longer improved.

In an early version of FALS, FALS 89 (*5*), only the correction term defined as Equation 6 was used throughout the iterative adaptation.

The results of FALS are validated by the leave-one-out prediction (*10*), which reconstructs the QSAR model by removing each compound once and predicts the removed compound. The discriminant function with a scientifically reasonable set of descriptors giving the best leave-one-out prediction should be finally adopted as the QSAR model.

Aquatic Toxicity of Organic Chemicals

The ranking of chemical substances with respect to their potential hazardous effects on aquatic life has become a subject of great importance in environmental sciences and toxicology. A number of quantitative structure-activity relationship (QSAR) studies have been reported using small congeneric series of organic compounds (for a recent review of the application of QSAR in aquatic toxicology, see (*11*)). However, it appears unlikely that the conclusions derived from the studies of only a few different classes of compounds are valid for most chemical pollutants. Recently, we formulated a quantitative structure-activity relationship (QSAR) model for the prediction of aquatic toxicity of organic chemicals (*12*). This section describes a study with toxicity data of 394 miscellaneous organic compounds using the FALS 89 (*5*) analysis.

Three hundred seventeen entries out of 394 were cited from "Water Quality Characteristics of Hazardous Materials," published by the National Technical Information Service (*13*). Toxicity is originally described in five grades, based on the "TLm96" test – the concentration of a substance that will kill 50% of the exposed test organisms within 96 h. The other 77 entries with the same toxicity description were found in the literature (*14–19*). Toxicity data were mainly on finfish, if available, and those on shrimp and other aquatic organisms were used to fill in the gaps. Because the number of compounds in the original five toxicity rating classes was unbalanced, we divided the compounds into three toxicity classes for analysis as follows:

Class 1 (practically nontoxic):	100 mg/L < TLm96
Class 2 (slightly toxic):	10 mg/L < TLm96 ≤ 100 mg/L
Class 3 (moderately or highly toxic):	TLm96 ≤ 10 mg/L

The typical structure types appearing in each class are alkanes, alkenes, and alcohols in Class 1, aromatic hydrocarbons and aromatic nitro compounds in Class 2, and phenols, organic phosphorus compounds, and aromatic halides in Class 3. Esters and aldehydes appeared in both Classes 1 and 2, and aliphatic halides in all classes. Three kinds of variables—continuous variables, discrete variables, and indicator variables—were used in the QSAR analysis. Molecular weight, hydrophobic constant (log P), and its squared value were used as continuous variables. The log P values included literature values (*20*) and some calculated values by using the CLOGP program (Pomona College MED CHEM Project, Claremont, CA). A discrete variable was defined as the number of specific atoms or constituents considered to contribute to toxicity. If the contribution was different in the electronic states of sp2 and sp3, in a ring and outside a ring, in an aliphatic moiety and in an aromatic moiety, then the number was counted in each case. Indicator variables were defined as one for the presence and zero for the absence of any kind of atom or moiety structure that contributed to toxicity, regardless of its number in a molecule. Some special descriptors were generated considering the substructures common to compounds which were misclassified in the analysis using general descriptors for atoms, bonds, functional groups, and physicochemical properties.

First, the relation of the continuous variables log P and molecular weight (MW) to the toxicity class was investigated using FALS89. The resultant correlations are given by equations 7–9.

$$\begin{aligned}Z &= 0.125 \log P - 0.869 \\ &\quad (CI = 0.25) \\ N &= 394 \quad MMG = 0.434 \quad N_{mis} = 217\ (15) \quad R_S = 0.314\ (p < 0.001)\end{aligned} \tag{7}$$

$$\begin{aligned}Z &= 0.708\ MW/100 - 1.399 \\ &\quad (CI = 0.59) \\ N &= 394 \quad MMG = 0.499 \quad N_{mis} = 195\ (9) \quad R_S = 0.410\ (p < 0.001)\end{aligned} \tag{8}$$

$$\begin{aligned}Z &= 0.061 \log P - 0.024(\log p)^2 + 0.704\ MW/100 - 1.442 \\ &\quad (CI = 0.12) \quad (CI = 0.24) \quad (CI = 0.58) \\ N &= 394 \quad MMG = 0.509 \quad N_{mis} = 195\ (7) \quad R_S = 0.447\ (p<0.001)\end{aligned} \tag{9}$$

where CI is the contribution index (= | coefficient | × standard deviation of the descriptor, a measure of the contribution of the descriptor to Z), N is the number of compounds, MMG is the mean membership grade, N_{mis} is the number of compounds misclassified, the figure in parentheses after N_{mis} is the number misclassified by two ratings, and R_S is the Spearman rank correlation coefficient. Although these correlations were statistically significant at the 0.1% level, the recognition results appeared rather poor. For the rate of uptake and nonspecific toxic action as narcosis, the most suitable parameter to describe acute toxicity is usually considered log P. However, the application of log P to the correlation study of heterogeneous compounds was not very effective.

We then added discrete and indicator variables for structural fragments and moieties, and examined these to formulate a good discriminant function. A satisfactory equation including 37 descriptors was derived at Fl = 0.1. The discriminant function (equation 10) is shown in Table I, where the descriptors with positive coefficients and those with negative coefficients are listed in the order of the CI values, indicating the degree of contribution to discrimination. Each substructural descriptor was independently counted. For example, 4-chlorophenol counted in both descriptors 1 and 10. Descriptors with positive coefficients are considered to contribute in a positive sense to the estimate of toxicity, whereas descriptors with negative coefficients contribute in a negative way. From the sign and contribution index for each variable, benzene and naphthalene rings, aliphatic and aromatic (more than one) halogen atoms, sulfur atoms, and so forth, apparently contribute to the enhancement of toxicity. Strictly speaking, however, these coefficients cannot be used to make general inferences about the contribution of each fragment within organic compounds as a broad class. They are valid only when used in the context of the present multidimensional model. MW and log P were not selected. Their contributions seem to be included in other descriptors appearing in the equation.

Table I. FALS QSAR Model[a] for Aquatic Toxicity (Equation 10)

	Descriptor[b,c]	*Coefficient*	*CI*
1	No. of benzene rings	0.710	0.39
2	No. of aliphatic halogens	0.232	0.30
3	ID for N(halogen in a benzene ring) ≥ 2	0.945	0.23
4	No. of sulfur atoms	0.647	0.22
5	ID for naphthalene ring	1.534	0.22
6	ID for heteroaromatic ring with Ad or halogen	1.514	0.19
7	ID for 7,7-dimethylbicyclo[2.2.1]hept-5-ene	1.345	0.18
8	No. of >P(=X)-X- (X: O,S)	0.532	0.15
9	ID for CH_2=CH-X (X: -CHO,-CH_2OH,-COOH,-CN)	1.428	0.14
10	ID for aromatic ring with A[d] and halogen	0.553	0.14
11	ID for -NO_2	0.510	0.11
12	ID for $X_1N(X_2)X_3$ [X_1-X_3: alkyl, N(C): 1-3]	0.674	0.10
13	ID for alcohol [N(C) in an alkyl ≥ 7]	0.795	0.09
14	No. of -OH, -SH (aromatic)	0.213	0.08
15	ID for 1,3-butadiene deriv.	0.818	0.07
16	$X_1C(X_2)(X_3)$-$C(X_4)(X_5)X_6$ [X_1-X_6: H or Cl; 3 for N(Cl) ≥ 4, 2 for N(Cl) = 3, 1 for N(Cl) ≤ 2]	0.299	0.07
17	ID for CH_3COOR [R: alkyl and N(C) in R: 4-5]	0.392	0.05
18	ID for -CHO	0.163	0.04
19	ID for acid anhydride	0.362	0.04
20	No. of -N< (aliphatic)	0.113	0.02
21	ID for cyclohexane	0.173	0.02
22	ID for three-membered ring	0.134	0.02
23	No. of -NH_2 (aliphatic)	0.049	0.01
24	ID for >N-N<	0.130	0.01
25	No. of heteroatoms in ring	0.006	0.01
26	ID for $X_1C(X_2)(X_3)X_4$ [X_1-X_4:halogen;N(Cl+Br) ≤ 2]	-1.017	0.12
27	ID for phenol without Ad at o-, p-, and with Bd	-0.773	0.10
28	ID for $X_1C(X_2)$=$C(X_3)X_4$ (X_1-X_4: H or halogen)	-0.754	0.09
29	$(\log P)^2$	-0.009	0.09
30	ID for $ClCH_2CH_2X$- (X: except >C<, >C=C<)	-0.858	0.09
31	ID for RCHO [R: alkyl, even N(C) in main chain]	-0.747	0.08
32	No. of -COOH	-0.163	0.07
33	ID for alkane, alkene	-0.232	0.06
34	No. of polar atoms (N, O)	-0.031	0.06
35	ID for alkyl [N(C): 4-5]	-0.021	0.01
36	No. of C (sp2, outside ring)	-0.005	0.01
37	No. of ketones	-0.022	0.01
	Constant	-0.685	

[a] Class boundaries: -0.289, 1.036.

[b] ID: Indicator variable. [c] N(x): No. of x. [d] A: Electron-donating substituents; B: Electron-accepting substituents.

(Reprinted with permission from Reference 12.)

Table II. Results of Recognition and Prediction Using 37 Descriptors

	Calculated Class		
Observed Class	1	2	3
Recognition[a]			
1	157	25	0
2	14	111	3
3	1	15	68
Leave-one-out prediction[b]			
1	155	27	0
2	21	102	5
3	3	21	60

[a] N=394, MMG=0.845, Nmis=58(1), Rs=0.857 ($p < 0.001$).

[b] N=394, MMG=0.802, Nmis=77(3), Rs=0.802 ($p < 0.001$).

(Reprinted with permission from Reference 12.)

The results of recognition by the discriminant function and prediction with the 37 descriptors are shown in Table II. The MMGs were fairly good, being 0.845 and 0.802, respectively, for recognition and prediction. The accuracy of classification into the three classes was 85.3% in the recognition and 80.5% in the leave-one-out prediction. Both may be considered satisfactory, despite the diversity of the molecular structures of organic chemicals investigated in this study.

Carcinogenicity in Rodents

The non-congeneric QSAR of carcinogenicity has become a subject of great importance for regulatory perspectives and ecotoxicity assessments. Approaches using some correlative methods such as the CASE fragment-based method and TOPKAT linear discriminant equation method were recently reviewed by Richard (*21*), who described that published prediction accuracies were in excess of 90%, while prospective prediction accuracies were less than 70% in those approaches. Moreover, worse results were published for a prospective prediction of rodent carcinogenicity using a variety of QSAR approaches (*22*). Further studies are required to improve the predictive reliability.

We have recently applied FALS 91 to a non-congeneric structure-carcinogenicity correlation using the data (Table III) (*23*) tested in rodent bioassays by the US National Toxicology Program (NTP). The test chemicals were segregated into six groups: Group A, carcinogenic to both rats and mice; Groups B–D, carcinogenic to only a single species at multiple sites, a single site in both sexes, and a single site in a single sex; Group E, giving only equivocal evidence of carcinogenicity; and Group F, non-carcinogenic. The data of Group E were excluded in our study. Inorganic and metalloorganic chemicals, polymers, and mixtures were also excluded from the analyses as shown in Table III. The database also included the results of Salmonella assays for mutagenicity. The FALS analyses were performed for carcinogenic / noncarcinogenic dichotomization using the following six sets of data assembled with consideration of the level and mechanism of carcinogenicity ; Set 1: A (75 compounds) / F (91 compounds), Set 2: A–D (155) / F (91), Set 3: (Salmonella positive) A (53) / F (24), Set 4: (Salmonella positive) A–D (87) / F (24), Set 5: (Salmonella negative) A (22) / F (64), and Set 6: (Salmonella negative) A–D (67) / F (64). Sets 1 and 2 are for overall carcinogenicity, Sets 3 and 4 for genotoxic carcinogenicity, and Sets 5 and 6 for non-genotoxic carcinogenicity.

Table III. Data for Non-Congeneric QSAR Analysis of Carcinogenicity

Group	*Carcinogenicity*	*No. in Group*	*Inorganics & Metallo-organ.*	*Polymers & Mixtures*	*Salmonella Assay*		
					Positive	*Negative*	*NT* ^{a}
A	both rats and mice	80	1^b	4^b	53	22	0
B-D	single species	82	0	2^b	34	45	1^c
	(B: multiple sites; C: single site, both sexes; D: single site, single sex)						
E	equivocal	39^b	4^b	0	11^b	24^b	0
F	non-carcinogenic	100	1^b	8^b	24	64	3^c

[a] Not tested. [b] Excluded. [c] Excluded from Sets 3–6.

Candidate descriptors were similar to those used in the analysis of aquatic toxicity. Continuous variables, discrete variables, and indicator variables for substructural and physicochemical features of molecules were investigated in the FALS analysis. The log P (octanol/water) values used were calculated by a simple method (*24, 25*) developed in our laboratory.

Table IV shows the QSAR model (equation 11) derived with Set 1 data consisting of Group A (carcinogenic to both rats and mice) and Group F (non-carcinogenic to both species). A set of 30 descriptors listed in the table were effective. N in the brackets after the descriptors indicates the numerical variable (continuous or discrete variable), and I in the brackets is the indicator variable. The descriptors with positive coefficients and those with negative coefficients are separately listed in the order of the contribution index. The boundary between Class 1 (non-carcinogenic compounds) and Class 2 (carcinogenic compounds) was 0.096. In the recognition, 13 chemicals among 166 were misclassified. The false positive is 6.0%, and the false negative is 1.8%. In the leave-one-out prediction, 34 chemicals among 166 were misclassified. The false positive is 11.4%, and the false negative is 9.0%. The results of leave-one-out prediction is somewhat lower than those of recognition, as a matter of course, but they are still considered satisfactory.

The results of all six sets are summarized in Table V. They are all considered satisfactory despite the diversity of the molecular structures of compounds investigated. It should be noted that the pre-classification by Salmonella assay markedly enhances the reliability, especially for false negatives in both recognition and leave-one-out prediction. However, the Salmonella test for the prediction of carcinogenicity has recently been underestimated.

Salmonella mutagenicity can be also predicted by the FALS approach. The results of FALS analysis of NTP mutagenicity data of 242 non-congeneric compounds are shown at the bottom of Table V, indicating that the reliability is better than those of carcinogenicity.

The preclassification based on mutagenicity is a dichotomization of genotoxic and non-genotoxic chemicals; it is considered as a kind of mechanism-based approach. It can be expected that much more information on the molecular mechanism of carcinogenicity will become available in the near future. Rational preclassification based on the molecular mechanism should be extensively investigated to enhance the prospective prediction accuracy of the non-congeneric QSAR approach for carcinogenicity.

QSARs on Biodegradation

Non-congeneric QSAR on biodegration has also become a subject of great importance for many authorities and industries responsible for hazard and risk assessment. Most of the published QSARs, however, are only valid for small groups of compounds of

Table IV. FALS QSAR Model[a] for Carcinogenicity from Set 1 (Equation 11)

	Descriptor[b]	*Coefficient*	*CI*
1	No. of H(-C)	0.052	0.53
2	No. of aromatic prim. amines	0.749	0.48
3	Proximity effect of N, O[c]	0.154	0.46
4	No. of saturated O (a,b-saturated)	0.293	0.32
5	No. of OH (aromatic & enols)	0.657	0.26
6	ID for rings except benzene & its condensed rings	0.500	0.25
7	No. of ureas	1.107	0.21
8	Weighted no. of hydrophobic atoms (C & Cl: 1.0, F: 0.5, Br: 1.5, I: 2.0)[c]	0.037	0.19
9	ID for basic moieties	0.375	0.18
10	Log P (calc.)[c]	0.095	0.15
11	ID for XC-CY (X,Y: N,O)	0.309	0.11
12	No. of COOH	0.239	0.11
13	ID for Ar-NH-R (R: except H)	0.299	0.10
14	ID for 2,4,5-unnsubstituted anilines	0.163	0.06
15	No. of -COO- (esters & lactones)	0.007	0.01
16	No. of sp^3C	-0.165	0.92
17	No. of H(-N, -O)	-0.350	0.58
18	No. of benzene rings	-0.385	0.29
19	No. of unsaturated N (in rings)	-0.722	0.26
20	No. of lactone rings	-1.463	0.25
21	No. of saturated S (a,b-saturated)	-1.065	0.24
22	ID for $R-SO_2NH-CO-R'$	-1.280	0.22
23	No. of unsaturated S	-0.408	0.19
24	No. of aliphatic tert. amines	-0.691	0.17
25	No. of intramolecular hydrogen bondsc	-1.230	0.16
26	ID for aromatic hydrocarbon multicycles	-0.660	0.15
27	No. of aliphatic ketones	-0.540	0.12
28	ID for 2 (or more) heteroatoms in a heteroarom. ring	-0.457	0.09
29	No. of unsaturated N (outside rings)	-0.071	0.05
30	No. of sp^2C (outside rings)	-0.035	0.03
	Constant	-0.574	

[a]Class boundary: 0.096. [b]ID: Indicator variable. [c]Data from Reference 24.

Table V. Results of FALS Analysis of Carcinogenicity and Mutagenicity

Set of data (Number of compounds)	*number of descriptors*	*Recognition* *false positive*	*Recognition* *false negative*	*Leave-one-out* *false positive*	*Leave-one-out* *false negative*
Carcinogenicity					
A (75) / F (91)	30	6.0%	1.8 %	11.4 %	9.0 %
A-D (155) / F (91)	34	6.1	4.9	13.4	8.9
Salm.Positive					
A (53) / F (24)	13	2.6	1.3	10.4	5.2
Salm.Positive					
A-D (87) / F (24)	15	4.5	0.9	10.8	3.6
Salm.Negative					
A (22) / F (64)	16	5.8	0.0	8.1	4.7
Salm.Negative					
A-D (67) / F (64)	22	9.9	1.5	13.0	7.6
Salm.Mutagenicity					
Pos. (113) / Neg. (131)	26	1.7	2.9	5.8	7.4

SOURCE: Adapted from ref. 9.

distinct chemical classes. We have recently applied FALS 91 to a non-congeneric structure-biodegradability correlation using the data of 1259 compounds collected from the database JETOC/KASHIN (*26*).

In the database, compounds were divided into several chemical groups specified by the guidelines of the Japanese Chemical Substances Control Law (JCSCL). We analysed Group 2 and Group 3 data; Group 2 data covered 463 aliphatic acyclic compounds (239 easily degradable and 224 not-easily degradable) having molecular weights of 41.1–971.2, and Group 3 data covered 796 aromatic monocarbocyclic compounds and their substituted derivatives (187 easily degradable and 609 not-easily degradable) having molecular weights of 68.1–1573.8. The degradation data of the database JETOC/KASHIN were obtained by the test procedure specified in the guidelines of JCSCL, usually based on the (BOD – B) / TOD value, where BOD is the oxygen demand from microorganisms during the test with the substance, B is that without the substance, and TOD is the theoretical oxygen demand for complete oxidation of the substance; easily degradable (Class 1) when the (BOD – B) / TOD ≥ 60%, otherwise not-easily degradable (Class 2).

Candidate descriptors included continuous variables, discrete variables, and indicator variables for substructural and physicochemical features of the molecules. Most of them were automatically generated using an extended chemical graph (*27*), which is a node connection table that adds abstracted nodes to the atom connection table of molecules. These descriptors were investigated in the FALS analyses.

Table VI shows the QSAR model (equation 12) derived with 463 aliphatic acyclic compounds. A set of 46 descriptors listed in Table VI were effective. With this model, 37 chemicals among 463 were misclassified in the recognition; 92.0% of the chemicals were correctly recognized. The result of the leave-one-out prediction was $N_{mis} = 68$ (85.3% correct).

For the aromatic monocarbocyclic compounds, a satisfactory QSAR model was obtained using 35 descriptors. Using this model, 53 of 796 chemicals were

Table VI. FALS QSAR Model[a] of Biodegradability for Aliphatic Acyclic Compounds (Equation 12)

Descriptor[b]	Coefficient	CI	Descriptor[b]	Coefficient	CI
1 SQRT(MW/100)	1.415	0.63	24 -CO-	-0.556	0.12
2 CC(4-6)[c] &[d] -CO-NH-	1.779	0.30	25 -COX	-0.590	0.12
3 CC(1-3)-O-C-O	1.351	0.22	26 -O-CO-O-	-1.340	0.11
4 CC(1-3)-NH_2	0.710	0.16	27 CC(4-6,UC>2[e])	-0.514	0.09
5 Cl	0.174	0.15	28 Epoxy	-0.266	0.08
6 -O-O-	0.528	0.14	29 CC(1-3)-O-O-	-0.527	0.08
7 CC(1-3) & -O-P	0.712	0.14	30 -CN	-0.301	0.08
8 S	0.192	0.13	31 CC(1-3) & -CO-NH-	-0.288	0.08
9 spC	0.279	0.11	32 CC(1-3) & -PO_4	-0.632	0.07
10 -CO-N<	0.805	0.11	33 -CHO	-0.539	0.07
11 -C=N-O	1.012	0.08	34 CC(>7)-O-	-0.572	0.07
12 -O-	0.052	0.06	35 CC(>7, Xn>2)	-1.413	0.07
13 -N-N-	0.413	0.06	36 -COO	-0.061	0.05
14 CC(1-3) & -CO-N-	0.177	0.06	37 X-COO-	-0.320	0.04
15 -P-O	0.266	0.05	38 CC(4-6)-O-	-0.180	0.04
16 -COO-	-0.437	0.43	39 CC(4-6) & =N-	-0.260	0.03
17 sp^{3}C	-0.051	0.35	40 AMP[f]	-0.446	0.03
18 CC(4-6) & -CO-N	-1.700	0.34	41 CC(>7)-N-	-0.182	0.03
19 CC(>7) & -CO-N<	-1.974	0.16	42 -CO-NH-	-0.055	0.03
20 O-C-O	-0.789	0.16	43 -CO-NH_2	-0.224	0.02
21 -NH2	-0.478	0.15	44 CC(>7)-OH	-0.238	0.02
22 CC(1-3) & O-S=O	-0.563	0.13	45 CC(4-6)-N-	-0.045	0.01
23 -PO-OH	-0.460	0.13	46 CC(4-6)-COOH	-0.045	0.01
			Constant	-1.252	

[a] Class boundary: 0.033. [b] Number of the specified substructures except SQRT(MW/100) and AMP; a terminal bond such as R-, -R, -R-, -R< etc. connects R with a carbon atom; -CO-N includes -CO-NH_2, -CO-NH-, and -CO-N<, but -CO-N- includes -CO-NH- and -CO-N<; X: a halogen atom. [c] Carbon chain having 4–6 carbon atoms. [d] CC(4-6) bound to -CO-NH- at an unspecified position. [e] More than 2 unsaturated carbon atoms included. [f] Amphoteric property
(Data from Reference 24.)

missclassified in the recognition; 93.3% were correctly recognized. The result of the leave-one-out prediction was also satisfactory; N_{mis} = 76 (90.5% correct). These results are fairly good despite the diversity of the molecular structures of the compounds investigated.

Toxicity measurements are based on results of specific and nonspecific multiple biological reactions. Any test end point may involve various combinations of toxic manifestations. Using QSAR in the toxicology of compounds having a wide variety of structures appears to be an extremely complicated problem, and the present studies using FALS may be far from complete. However, such approaches will become important as more complex molecules with multiple functions are developed and made commercially available. Recently, regulatory agencies and industry have attempted to use the non-congeneric QSAR to make rapid and cost-effective predictions about the toxicity of industrial chemicals. We expect that non-congeneric QSAR studies like ours would greatly enhance the capabilities of regulators and producers to screen compounds for possible environmental risks and pollutant hazards.

Literature Cited

1. Moriguchi, I.; Komatsu, K. *Chem. Pharm. Bull.,* **1977**, *25*, 2800 .
2. Moriguchi, I.; Komatsu, K.; Matsushita, Y. *J. Med. Chem.*, **1980**, *23*, 20 .
3. Moriguchi, I.; Komatsu, K. In *Abstracts of Papers, 8th Symposium on Structure-Activity Relationships*; Tokyo, Oct., **1981**; pp 55–58.
4. Novak, V. In *Fuzzy Sets and Their Applications*; Adam Hilger: Bristol, **1989**; Chapter 9.
5. Moriguchi, I.; Hirono, S.; Liu, Q.; Matsushita, Y.; Nakagawa, T. *Chem. Pharm. Bull.,* **1990**, *38*, 3373 .
6. Moriguchi, I.; Liu, Q.; Hirono, S.; Matsushita, Y.; Nakagome, I. In *Abstracts of Papers, 19th Symposium on Structure Activity Relationships*; Kawaguchi, Nov., **1991**; pp 244–247.
7. Moriguchi, I.; Hirono, S.; Matsushita, Y.; Liu, Q.; Nakagome, I. *Chem. Pharm. Bull.*, **1992**, *40*, 930 .
8. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I. *Quant. Struct.-Act. Relat.*, **1992**, *11*, 325 .
9. Moriguchi, I. *Yakugaku Zasshi*, **1994**, *114*, 135.
10. Tukey, J. W. *Ann. Math. Stat.*, **1958**, *29*, 614 .
11. Hermens, J. L. M. *Pestic. Sci.*, **1986**, *17*, 287.
12. Liu, Q.; Hirono, S.; Matsushita, Y.; Moriguchi, I. *Environ. Toxicol. Chem.*, **1992**, *11*, 953.
13. Hann, R. W.; Jensen, P. A. In *Water Quality Characteristics of Hazardous Materials*; PB-285946; National Technical Information Service: Springfield, VA., **1977**.
14. Hodson, P. V.; Dixon, D. G.; Kaiser, K. L. E. *Environ. Toxicol. Chem.*, **1984**, *3*, 243.
15. Mortimer, J. K. ; Ruth, M. D. *Environ. Sci. Technol.* , **1987**, *21*, 149.
16. Zaroogian, G.; Heltshe, J. F.; Johnson, M. W. *Aquat. Toxicol.* , **1985**, *6*, 251.
17. Thurston, R. V.; Gilfoil, T. A.; Meyn, E. L.; Zajdel, R. K.; Aoki, T.I.; Veith, G. D.*Water Res.* , **1985**, *19*, 1145.
18. Finlayson, B. J.; Verrue, K. M. *Arch. Environ. Contam. Toxicol.*, **1985**, *14*, 153.
19. Neely, W. B. *Chemosphere*, **1984**, *13*, 813.
20. Hansch, C.; Leo, A. In *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, NY., **1979**.
21. Richard, A. M. *Mutation Res.*, **1994**, *305*, 73.
22. Hileman, B. *Chem. Eng. News*, **1993**, *71*, No.25, 35.
23. Ashby, J.; Tennant, R. W. *Mutation Res.*, **1991**, *257*, 229.
24. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. *Chem. Pharm. Bull.*, **1992**, *40*, 127 .
25. Moriguchi, I.; Hirono, S.; Nakagome, I.; Hirano, H. *Chem. Pharm. Bull.,* **1994**, *42*, 976 .
26. The data were kindly provided by Dr. T. Oshima, Japan Chemical Industry Ecology-Toxicology and Information Center, 1-19-4 Nishi Shinbashi, Minato-ku,Tokyo 105, Japan.
27. Hirano, H.; Nakagome, I.; Hirono, S.; Moriguchi, I. In *Abstracts of Papers, 21st Symposium on Structure-Activity Relationships*; Tokushima, Nov., **1993**; pp 189–192.

RECEIVED June 5, 1995

Chapter 12

Computer-Aided Molecular Modeling and Structure–Activity Analyses of New Antifungal Tertiary Amines

Chiyozo Takayama, Naoto Meki[1], Yasuyuki Kurita, and Hirotaka Takano

Agricultural Chemicals Research Laboratory, Sumitomo Chemical Company, Ltd., 4–2–1, Takatsukasa, Takarazuka 665, Japan

Various computational procedures such as the Hansch-Fujita method and the ACACS molecular graphics system have been used in agrochemical design research in our laboratory for many years. The ACACS system was applied to the structure-activity analyses of antifungal amine compounds. We designed two series of azasterol mimics on the basis of the structure of fecosterol, the substrate of the $\Delta^8 \rightarrow \Delta^7$-isomerization step in the ergosterol biosynthetic pathway of fungi. As expected, two types of new tertiary amines, which we synthesized, showed a high antifungal activity against various plant-pathogenic fungi. One of the series inhibited well the $\Delta^8 \rightarrow \Delta^7$-isomerase as expected, while the other blocked strongly the Δ^{14}-reduction step, but scarcely the $\Delta^8 \rightarrow \Delta^7$-isomerization. The structure-activity relationships were reasonably rationalized by conformational analyses using the MNDO-PM3 molecular orbital method in the MOPAC program package incorporated in the ACACS system.

Various computational procedures are used in designing new agrochemicals and medicines these days. In our laboratory, the classical Hansch-Fujita QSAR (Quantitative Structure-Activity Relationship) method (*1*) has been used in agrochemical design research for many years. We have also been using a number of other procedures such as theoretical physicochemical calculations and molecular modeling. Various computer programs and databases have been organized so that they are interactive with each other.

In this chapter, after a brief introduction to computational systems used in the Sumitomo Chemical, an example of the application to structure-activity analyses of new antifungal tertiary amines is described.

[1] Current address: Petrochemicals Research Laboratory, Sumitomo Chemical Company, Ltd., 2–1, Kitasode, Sodegaura 299–02, Japan

0097–6156/95/0606–0154$12.00/0

Computer-Aided Agrochemical Research

Computational procedures used for agrochemical research in our laboratory are shown in Figure 1. Not only classical and modern QSAR methods but also various statistics, molecular graphics, and knowledge base systems are included. In these systems, databases of 2-D (two-dimensional) and 3-D molecular structures and biological activities are incorporated. Furthermore, theoretical and empirical calculations are performed for data generation.

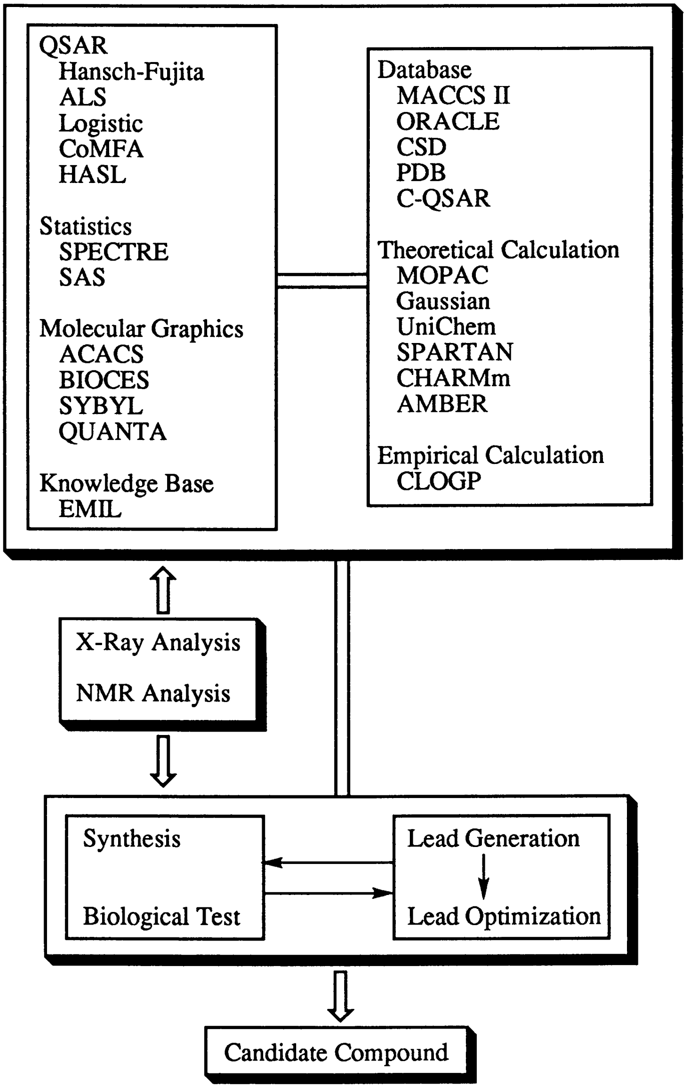

QSAR
Hansch-Fujita
ALS
Logistic
CoMFA
HASL
Statistics
SPECTRE
SAS
Molecular Graphics
ACACS
BIOCES
SYBYL
QUANTA
Knowledge Base
EMIL
Database
MACCS II
ORACLE
CSD
PDB
C-QSAR
Theoretical Calculation
MOPAC
Gaussian
UniChem
SPARTAN
CHARMm
AMBER
Empirical Calculation
CLOGP
X-Ray Analysis
NMR Analysis
Synthesis
Biological Test
Lead Generation
Lead Optimization
Candidate Compound

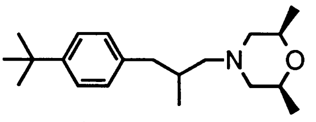
Figure 1. Computer-aided agrochemical design systems in the Sumitomo Chemical.

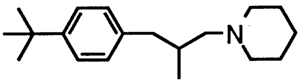
In Figure 1, the SPECTRE (Statistical Predictive Expertizing in Chemical and Technological Refined Experimentation) (*2*) is a chemometrics system co-developed by Professor Mario Marsili of University of L'Aquila, Italy and the Sumitomo Chemical. In this system, the PLS (Partial Least Squares) regression method (*3*) is employed for data analyses. The ACACS (Advanced Computer Aided Chemistry System) (*4, 5*) and BIOCES (Biochemical Expert System) (*6*) systems have been developed through cooperations with some other industries. The ACACS is potentially applicable for modeling small molecules such as agrochemicals and medicines. The BIOCES is mainly for the protein modeling. The EMIL (Example Mediated Innovation for Lead Evolution) system is for the database aided lead evolution of bioactive compounds (*7*), which has been constructed by Professor Toshio Fujita of Kyoto University with his coworkers. The MACCS II (*8*) and ORACLE (*9*) systems are used for the storage and retrieval of chemical structures and biological activities of a vast number of compounds tested in this laboratory. The databases of the QSAR correlation equations, molecular hydrophobicity log P values, and other physicochemical parameters are organized in the C-QSAR system developed by Professors Corwin Hansch and Albert Leo of Pomona College and their coworkers (*10*).

We have been using the Hansch-Fujita method effectively in designed syntheses and developmental studies of amide and urea herbicides (*11–15*), and imide and carbamate fungicides (*16–19*). This method has been shown to be in fact a powerful tool for optimizing a given lead structure as well as for identifying the "best" compounds, for example, in developing bromobutide, a broad spectrum *N*-benzylacylamide herbicide, which is marketed in Japan at present (*20*). The method is useful also in decision-making for termination of projects for series syntheses (*16*). The ALS (Adaptive Least Squares) method is a variation of QSAR to analyze activity data represented by the rating scores (*17, 18, 21*). The ACACS system has been used in molecular modeling studies of various agrochemicals such as pyrethroids and GABA (γ-aminobutyric acid) inhibitory insecticides (*22, 23*). In these ACACS studies, similarities in 3-D structures among compounds are examined on the basis of conformational analyses with molecular orbital methods. We have applied the ACACS to rationalize the structure-activity relationship of novel antifungal tertiary amines inhibiting ergosterol biosynthesis of plant-pathogenic fungi. The details will be described below.

New Antifungal Amine Compounds

Design and Synthesis: Ergosterol plays an essential role in the structure and function of the fungal cell membrane. Among various fungicides, whose intrinsic action is to inhibit the ergosterol biosynthesis, so-called azole compounds such as diniconazole-M and propiconazole inhibit the 14α-demethylation step in the ergosterol biosynthesis pathway (*24*). The target steps of amine fungicides such as fenpropimorph and fenpropidin are, however, the sterol Δ^{14}-reductase and /or $\Delta^{8}\rightarrow\Delta^{7}$-isomerase reactions as shown in Figure 2 (*25, 26*).


fenpropimorph


fenpropidin

Certain azasterols inhibit particular enzymes catalyzing some structural conversions of sterols occurring in the vicinity of the position corresponding to the

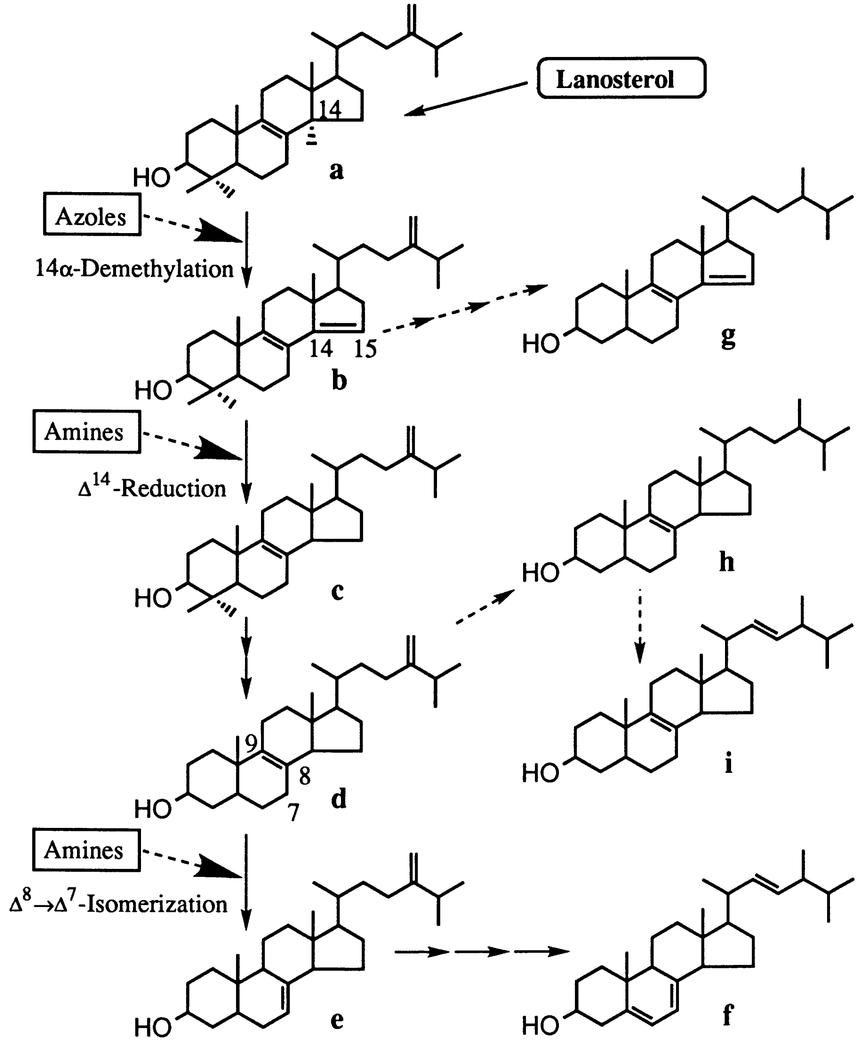
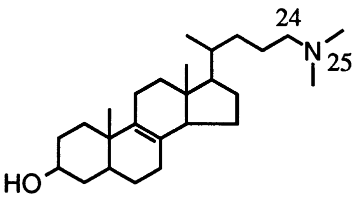

Lanosterol
14
HO
a
Azoles
14α-Demethylation
14 15
b
HO
g
Amines
Δ14-Reduction
HO
c
HO
h
9
8
7
d
HO
i
Amines
Δ8→Δ7-Isomerization
HO
e
HO
f

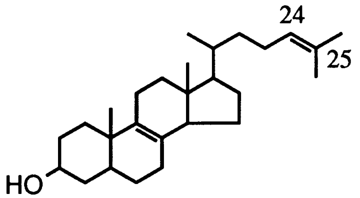
Figure 2. Ergosterol biosynthesis pathway from lanosterol and sites of inhibition of azole and amine fungicides. **a**, 24-methylene-dihydrolanosterol; **b**, 4,4-dimethylergosta-8,14,24(28)-trienol; **c**, 4,4-dimethylergosta-8,24(28)-dienol; **d**, fecosterol; **e**, episterol; **f**, ergosterol; **g**, ignosterol; **h**, ergost-8-enol; **i**, ergosta-8,22-dienol.

introduced nitrogen atom (*27, 28*). For example, 25-aza-24,25-dihydrozymosterol inhibits the Δ^{24}-sterol methyltransferase of *Saccharomyces cerevisiae* which catalyzes $\Delta^{24(28)}$-methylene generation as in the conversion of zymosterol to fecosterol (*28*).

About ten years ago, we started a synthetic project for amine compounds to obtain new fungicides. We selected the $\Delta^8 \rightarrow \Delta^7$-isomerization step of fecosterol to



25-aza-24,25-dihydrozymosterol



zymosterol

episterol as the target and applied the above information about the aza-introduction in the course of the molecular design. Two series of azasterol mimics **I** and **II** were designed on the basis of the structure of fecosterol, the substrate of the isomerization enzyme (Figure 3). In the structures **I** and **II**, the nitrogen atom is placed at the location corresponding to the position 8 and 9 of the sterol skeleton, respectively. In fact, compounds **1** and **2** (Figure 3) belonging to structures **I** and **II**, where the C-17 side chain of the steroid skeleton is imitated by *p*-Cl-phenoxyphenyl and *t*-Bu-phenyl, respectively, were synthesized. They showed a high antifungal activity against various plant-pathogenic fungi (Table I). A number of amines of the compound **1**-type such as those shown in Figure 4 were synthesized. Table II shows that compounds **1**, **3**, and **4** are active, but compound **5** is inactive curatively against powdery mildew of wheats in the pot test. The antifungal activity of compounds **1** and **2** seems to be slightly lower than that of fenpropimorph against *Ustirago maydis*, *in vitro*, as shown in Table III.

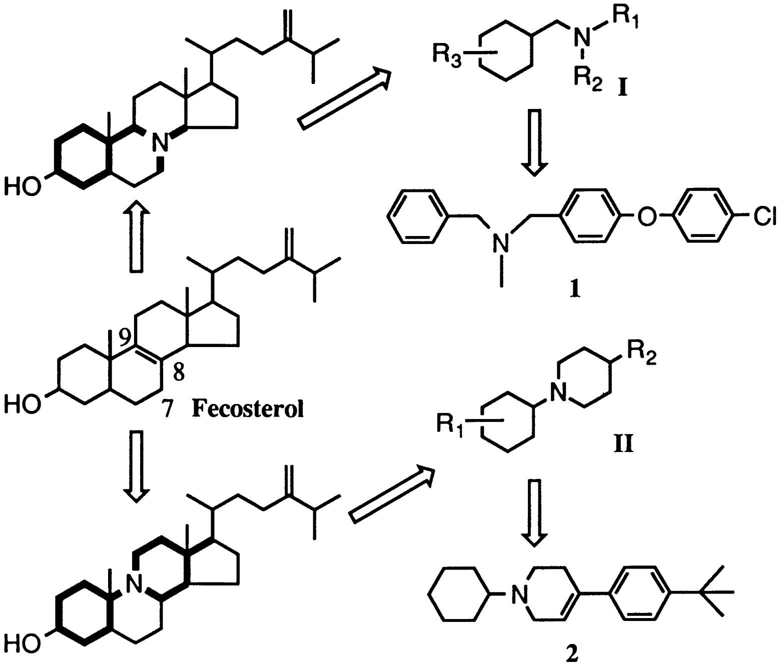


Figure 3. Design of ergosterol biosynthesis inhibitors.

Table I. Antifungal Activities *In Vitro* of Compounds 1 and 2

Compound	EC_{50}[a]					
	U.n.[b]	*A.k.*[c]	*C.m.*[d]	*C.l.*[e]	*B.c.*[f]	*P.o.*[g]
1	2.0	4.8	1.0	1.5	5.9	2.5
2	1.3	1.9	0.4	1.0	1.2	1.0

[a] Concentration (mg/l) required to give 50% inhibition of mycelial growth. The value was measured according to the procedure described in Kato, T.; Tanaka, S.; Yamamoto, S.; Kawase, Y.; Ueda, M. *Ann. Phytopathol. Soc. Jpn.* **1975,** *41,* 1. [b] *Ustilago nuda.* [c] *Alternaria kikuchiana.* [d] *Cochliobolus miyabeanus.* [e] *Colletotrichum lagenarium.* [f] *Botrytis cinerea.* [g] *Pyricularia oryzae.*

Sterol Analysis: To ascertain the target sites of newly synthesized compounds **1** and **2**, the sterols existing in *U. maydis* treated with amines were analyzed with gas chromatography-mass spectrometry (GC-MS) as previously described (*29*). Sporidia of *U. maydis* were grown for 24 hr at 27°C with and without the test compounds. The sterols being accumulated in cultures were then extracted and analyzed by the GC-MS analysis at 70 eV. Table III shows the relative amount of identified sterols in the extract.

The accumulation of ignosterol and that of ergost-8-enol and ergosta-8,22-dienol suggest, respectively, that the Δ^{14}-reductase and the $\Delta^{8}\rightarrow\Delta^{7}$-isomerase are inhibited (Figure 2) (*30*). Therefore, the results shown in Table III are considered to indicate the followings. (1) Compound **1** inhibits both steps, the susceptibility of the $\Delta^{8}\rightarrow\Delta^{7}$-isomerase being higher than that of the Δ^{14}-reductase. The relative amount of ignosterol accumulated by the treatment of 10 ppm of compound **1** is high, whereas ignosterol is not detected with 1 ppm. Because the Δ^{14}-reduction step occurs prior to the $\Delta^{8}\rightarrow\Delta^{7}$-isomerization, the ignosterol accumulation is higher than that of ergost-8-enol and ergosta-8,22-dienol when the reductase is inhibited first with the higher concentration of compound **1**. (2) Compound **2** blocks strongly the Δ^{14}-reduction, but scarcely the $\Delta^{8}\rightarrow\Delta^{7}$-isomerization.

Estimation of pK_a Values: Both of the Δ^{14}-reductase and $\Delta^{8}\rightarrow\Delta^{7}$-isomerase reactions are considered to proceed through high energy carbocationic intermediates

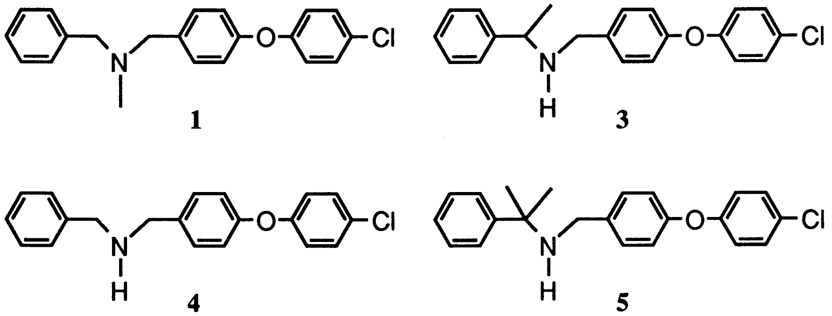

1 **3** **4** **5**

Figure 4. Structure of compound 1 and its analogs.

Table II. Curative Activity of Compound 1 and Its Analogs against Powdery Mildew of Wheats

	Conc. of the compound treated (mg/l)		
Compound	500	50	12.5
1	5[a]	5	4
3	5	5	4
4	5	4	2
5	0	0	0

[a]Disease control: 5, 100% control of infection; 4, 90–99%; 3, 70–89%; 2, 50–69%; 1, 30–49%; 0, 0–29%. The activity was measured according to the procedure described in Takano, H.; Oguri, Y.; Kato, T. *J. Pestic. Sci.* **1986,** *11,* 373.

shown in Figure 5 (*25*). These intermediates have a positive charge formally at the C-14 and at the C-8, respectively. Amine fungicides such as fenpropimorph and fenpropidin in the *N*-protonated forms are suggested to mimic the carbocationic intermediates in the reduction as well as the isomerization (*25, 31*). Their pK_a values described in literatures are *c.* 7.5 (*32*) and *c.* 10.5 (*33*), respectively. From these values, these two fungicides would exist mostly in their protonated form under the physiological pH conditions (*25, 31*).

To judge whether amines **1** to **5** are protonated under the physiological pH or not, we estimated their pK_a values using empirical correlation equations 1 and 2 for aliphatic amines ($NR_1R_2R_3$) proposed by Takayama *et al.* (*34*) and Hall (*35*), respectively. Equation 1 is for the primary and secondary amines, while equation 2 is for the tertiary amines.

Table III. Relative Amount of Identified Sterols in Cultures of *Ustirago Maydis* Treated with Amines

Treatment		Sterols (%)				
Compound	Conc. (mg/l)	IST[a]	E-8[b]	E-8,22[c]	EST[d]	Others
Control		n.d.[e]	n.d.	n.d.	68	32
1	10	44	17	35	n.d.	4
1	1	n.d.	62	36	n.d.	2
2	10	96	n.d.	3	n.d.	1
2	1	90	n.d.	n.d.	9	1
2	0.1	50	n.d.	n.d.	27	23
Fenp[f]	1	88	n.d.	11	n.d.	1
Fenp	0.1	n.d.	44	42	n.d.	14

[a]Ignosterol. [b]Ergost-8-enol. [c]Ergosta-8,22-dienol. [d]Ergosterol. [e]Not detected. [f]Fenpropimorph.

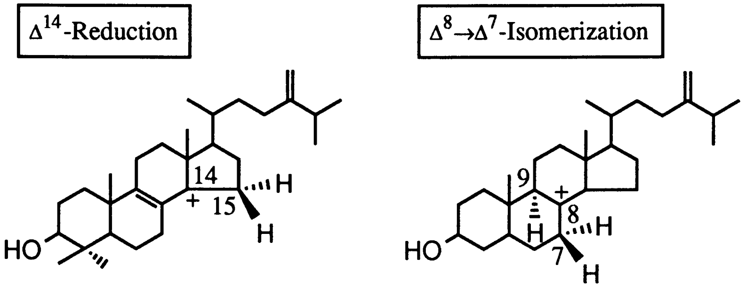


Figure 5. Structure of the high-energy carbocationic intermediates involved in the sterol Δ^{14}-reductase and $\Delta^{8}\rightarrow\Delta^{7}$-isomerase reactions.

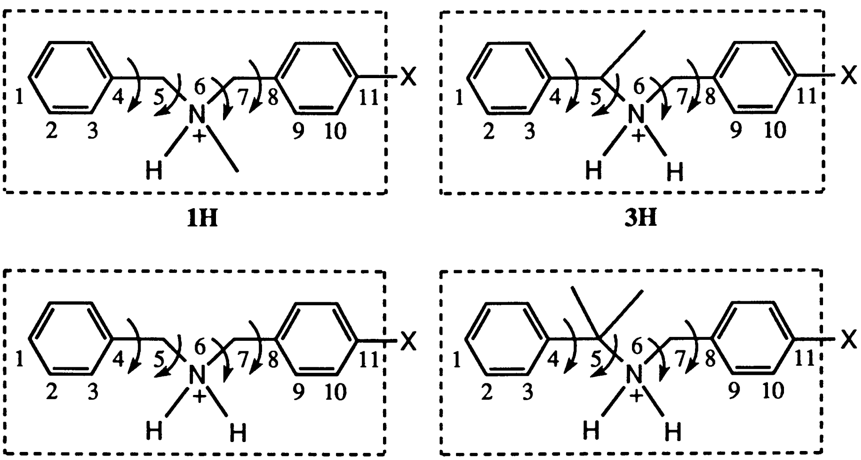
$$pK_a = -3.603\Sigma\sigma^* + 0.335[E_s^c(R_2) + E_s^c(R_3)] + 1.427n_H + 9.545 \qquad (1)$$

$$pK_a = -3.30\Sigma\sigma^* + 9.61 \qquad (2)$$

In these equations, $\Sigma\sigma^*$ is the summation of the Taft σ^* values (*36*) of three N substituents. $E_s^c(R_2)$ and $E_s^c(R_3)$ are, respectively, the Hancock corrected steric constants (*37*) of two bulkier substituents R_2 and R_3 among three N substituents, in which R_1 = H and $E_s^c(R_2) \geq E_s^c(R_3)$, and n_H is the number of hydrogen atoms in respective conjugate ammonium ions. σ^* and E_s^c values used in the calculations were taken from literature (*34*, *36*, *38*). For the 4-(4-chlorophenoxy)benzyl group in amines **1**, **3**, **4**, and **5**, σ^* and E_s^c values of benzyl were used (*39*). The σ^* and E_s^c values of the bidentate ligand in the amine **2** were taken as the sum of σ^*(Et) and σ^*(Allyl), and twice E_s^c(Et), respectively (*34*). The σ^* value of the α,α-dimethylbenzyl group in the amine **5** was estimated by using those of Me, Et, and benzyl groups on the basis of the additive nature of σ^* values (*36*, *40*). Its E_s^c value was taken as that of *t*-Pent group. The calculation with equations 1 and 2 estimated the pK_a value as follows: **1**, 8.19; **2**, 9.94; **3**, 8.63; **4**, 8.62; **5**, 8.25. Thus, all the 5 amines are reasonably considered to exist mostly in their protonated forms under the physiological pH conditions as in the case of fenpropimorph and fenpropidin.

Molecular Orbital Calculations: To examine the assumption that the amine fungicides mimic the carbocationic intermediates of sterols in the targeted enzymic reactions stereoelectronically, conformational analyses by using the MNDO-PM3 molecular orbital method (*41*) incorporated in the MOPAC (Molecular Orbital Package) version 7 (1993) (*42*) were made for the two carbocationic intermediates and the protonated form of compounds **1** to **5**. The initial geometries for the calculations were generated from standard bond lengths and angles using the Model Builder program integrated into the ACACS system, unless otherwise noted.

Active Conformation: To generate the initial coordinates of the 3-D structure of the carbocationic intermediates for the Δ^{14}-reduction and $\Delta^{8}\rightarrow\Delta^{7}$-isomerization, a rigid steroid skeleton was constructed by using the X-ray analysis data of a lanosterol derivative, 24,25-dibromolanost-8-en-3β-yl acetate (*43*). The conformation of the flexible side-chain at position 17 was taken to be fully extended and staggered. Jensen *et al.* have proposed this type of the initial conformation for the side-chain of



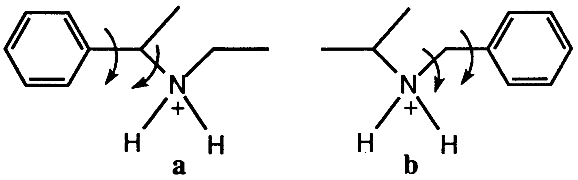
1H 3H

4H 5H

Figure 6. Structure of ammonium ions **1H**, **3H**, **4H**, and **5H** and their core substructure. X: O-C_6H_4-*p*-Cl. The substructure **1H'**, **3H'**, **4H'**, and **5H'** is enclosed with dotted lines.

the carbocationic intermediate in the $\Delta^8 \rightarrow \Delta^7$-isomerase reaction (*44*). The initial geometries were fully optimized by the MO calculation to obtain the minimum-energy conformation for each cationic intermediate. The optimized structures were adopted as candidates for the active conformation.

Although the target sites of action of compounds **3** and **4** were not examined, we considered that they share the same sites as compound **1** because of their closely related structures. Thus, comparative conformational analyses were carried out by taking the inactive analog compound **5** as the negative reference. To save computation time, the analyses were made for the core substructure of the protonated compounds **1**, **3**, **4**, and **5** (designated as **1H'**, **3H'**, **4H'**, and **5H'**, respectively) enclosed with dotted lines in Figure 6. Such a treatment is reasonable since the substituents omitted from the calculations are identical among all of the 4 compounds and they are at the *para* position. The 4-chlorophenoxy substituent does not seem to show significant effects on the conformation of the substructures. The ammonium ion **3H** has an asymmetric center at one of the benzyl carbon atoms, which of the enantiomers is biologically more active has not been investigated. The ammonium ion **1H** is chiral as the protonated form, although its enantiomers are interchangeable through the non-protonated form. Both enantiomeric configurations were considered in the conformational analyses for **1H'** and **3H'**. For each substructure, the geometry generated by the Model Builder program was optimized and then 81 (3x3x3x3) starting geometries were produced by rotating the bonds in the optimized structure, shown with arrows in Figure 6, by 0°, 120°, and −120°. These geometries were fully optimized to obtain stable conformers. Because no guarantee existed as to whether all of the low-energy minima are gleaned by such calculations (*45*), the following operations were additionally performed. For substructures **1H'**, **4H'**, and **5H'**, their symmetrical property was used to obtain conformers overlooked in the above analyses. The substructure **3H'** was further "divided" into two segments, **a** and **b**, as depicted in Figure 7. For these segments, the starting geometries were produced by changing each of the torsional angles indicated by arrows in steps of 30° and fully optimized. The stable conformers for segments were combined to reconstruct the substructure **3H'**. Its geometries were fully optimized to glean additional low-energy conformers for each enantiomeric substructure. With the above-mentioned operations, 31, 24, 19, and 21 stable conformers were obtained for **1H'**, **3H'**, **4H'**, and **5H'**, respectively. The torsional angles and conformational energies are listed in Table IV for the conformers of **1H'** having the (*S*) configuration. The signs of torsional angles are to be reversed for the (*R*) configuration.



a **b**

Figure 7. Segments of the substructure **3H'**.

Based on the above analyses, the shapes of the substructures **1H'**, **3H'**, **4H'**, and **5H'** were compared with the least square fitting method (*46*). The conformers of the substructures were fitted to each other at 7 points of C1, C4, C5, N6, C7, C8, and C11 shown in Figure 6. Assuming that the overall shape is important for the activity of these antifungal amine compounds, we set up the following criteria in evaluation of the shape similarity: the root mean square (RMS) value of distances between the atoms used to fit the conformers should be within 0.5 Å, and the difference in the dihedral angle of corresponding benzene rings between substructures should be within 30°. Similar criteria have been used to examine the shape similarity among insecticidal pyrethroids in our previous work (*22*).

When a compound is only weakly bioactive, the weak activity could be attributed to either a low affinity including a low steric complementarity for the target receptor or enzyme or a low concentration around the target site (*22*). The log *P* values of the compounds under consideration were calculated, with the CLOGP program (*47*) release 3.55 incorporated in the C-QSAR system 1.87 package (1994), to give the value of 6.18, 6.08, 5.77, and 6.45, respectively, to compounds **1**, **3**, **4**, and **5** (and 6.27 to compound **2**). Variations in the log *P* value are not so large. Furthermore, the pK_a value estimated above does not vary much among compounds **1**, **3**, **4**, and **5**. Hence, the inactivity of compound **5** was expected to be due mainly to an unfavorable 3-D structure in the protonated form and not to the transport process to the target site. In other words, the ammonium ion **5H** may not be able to take the conformation required for the high activity in the low energy conformational space.

The conformers of **1H'**, **3H'**, and **4H'** were compared with those of **5H'**. Seventeen among 31 low-energy conformers generated for **1H'** were judged to be dissimilar to any stable conformers of **5H'** (Table IV) on the basis of criteria for the shape similarity mentioned above. For **3H'** and **4H'**, 9 and 8 conformers dissimilar to those of **5H'** were found out, respectively. These results were taken to indicate that the active conformers of **1H'**, **3H'**, and **4H'** are among the 17, 9, and 8 conformers, respectively. This process is similar to that in which we pointed out previously a candidate for the active conformer of esfenvalerate, a highly active pyrethroid, based on the shape comparison with low activity compounds (*22*). Grdadolnik *et al.* also compared conformations of active and inactive backbone-cyclized analogs of substance P to elucidate features of the bioactive conformation (*48*).

The next step was to compare the shape of conformers among **1H'**, **3H'**, and **4H'** which were selected above by the comparison with those of **5H'**. For **1H'**, conformers 9, 10, 14, 15, 17, and 18 in Table IV were found to be similar to some of the low-energy conformers of **3H'** as well as **4H'**. The comparison of (*S*)-**1H'** with

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Table IV. Conformers of the Substructure 1H'

	Torsional Angles[b]				
Conformer[a]	τ_1	τ_2	τ_3	τ_4	ΔE[c]
1	104	84	–160	–100	0.0
2	100	160	–84	–105	0.0
3*	–80	82	–153	–77	0.0
4	77	153	–82	80	0.0
5*	100	161	–76	–81	0.2
6	81	76	–161	–100	0.2
7*	80	75	–153	106	0.3
8*	–106	153	–75	–80	0.3
9*	86	77	66	–94	0.6
10*	94	–66	–77	–86	0.6
11	99	161	–161	–99	0.6
12*	–109	149	–161	–99	0.6
13	99	161	–149	110	0.6
14*	–105	–77	–86	76	1.3
15*	–77	86	77	106	1.3
16	83	80	78	–77	1.5
17*	77	–78	–81	–84	1.5
18*	–73	87	69	–95	1.7
19	95	–69	–87	73	1.7
20*	–106	150	46	75	1.8
21	–76	–46	–150	107	1.8
22	–76	–46	–158	–101	1.9
23*	101	158	46	75	1.9
24*	–110	144	57	101	2.0
25	–101	–57	–145	110	2.0
26*	100	161	62	–92	2.0
27	92	–61	–160	–100	2.0
28	83	–60	–136	–94	2.1
29*	94	137	60	–83	2.2
30	101	171	80	116	2.7
31*	–115	–80	–171	–101	2.7

[a] The asterisks (*) refer to conformers which are judged to be dissimilar to any stable conformer of **5H'**. [b] Torsional angles in degree are defined with notations in Figure 6 as follows: τ_1, C3–C4–C5–N6; τ_2, C4–C5–N6–C7; τ_3, C5–N6–C7–C8; τ_4, N6–C7–C8–C9. [c] The difference in the heat of formation (kcal/mole) from that of the most stable conformer.

(*R*)-**3H'** and **4H'** [or that of (*R*)-**1H'** with (*S*)-**3H'** and **4H'**] pointed out conformers 9, 10, 15, and 18, and that of (*S*)-**1H'** with (*S*)-**3H'** and **4H'** [or that of (*R*)-**1H'** with (*R*)-**3H'** and **4H'**], conformers 9, 10, 14, and 17 to be similar. Among the above 6

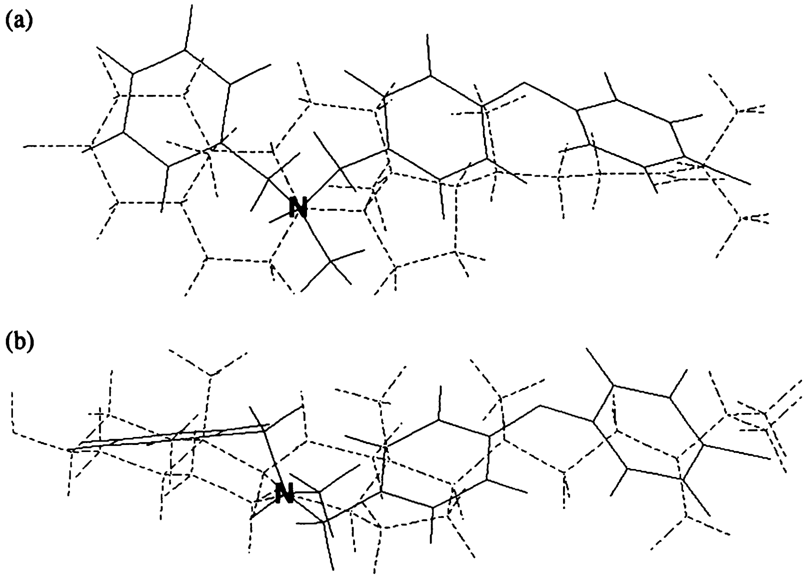
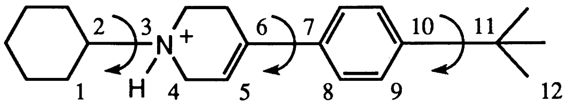


Figure 8. A candidate for the active conformer of the ammonium ion **1H** (solid line) and its superimposition on the carbocationic intermediate (dotted line) of the $\Delta^8 \rightarrow \Delta^7$-isomerase reaction. (a) and (b) are front and side views, respectively.

conformers, conformers 9 and 10 are energetically most favorable. They were adopted here as candidates for the active conformer of **1H'**. Candidates for the active conformer of the ammonium ion **1H** were constructed by attaching the *p*-chlorophenoxy group to conformers 9 and 10 of **1H'** and then fully optimized. The optimized structure derived from conformer 9 of **1H'** is shown in Figure 8.

For compound **2**, MO calculations were made using the ammonium ion **2H** (Figure 9). For the initial geometry, the hetero-atomic ring was attached to the chair-form cyclohexane ring equatorially. The structure of this ammonium ion is relatively rigid and almost symmetrical, although the N atom is an asymmetric center in the protonated form. Therefore, we thought that the number of the conformers in the low-energy conformational space is not so many and that the overall shape of the conformers is similar to each other. Starting geometries for the calculations were generated by changing each of the torsional angles indicated with arrows in Figure 9 in steps of 60° for the (*S*) configuration and fully optimized to obtain 12 conformers shown in Table V. The 3-D structure of the most stable conformer among those energetically minimized shown in Figure 10 was adopted as the candidate for the active conformer.



2H

Figure 9. Structure of the ammonium ion **2H**.

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Table V. Conformers of Ammonium Ion 2H Having (*S*) Configuration

Conformer	Torsional Angles[a]			ΔE[b]
	τ_1	τ_2	τ_3	
1	180	–144	89	0.0
2	64	–144	–89	0.1
3	180	–144	–91	0.2
4	–66	–144	89	0.5
5	–66	–144	–89	0.5
6	180	131	–87	0.5
7	180	132	89	0.5
8	64	131	89	0.7
9	64	131	–89	0.7
10	61	–144	89	0.7
11	–66	132	89	1.0
12	–67	131	–89	1.1

[a] Torsional angles in degree are defined with notations in Figure 9 as follows: τ_1, C1–C2–N3–C4; τ_2, C5–C6–C7–C8; τ_3, C9–C10–C11–C12. [b] The difference in the heat of formation (kcal/mole) from that of the most stable conformer.

Superimposition of 3-D Structures: Figure 8 shows the superimposition of the 3-D structure of the ammonium ion **1H** derived from conformer 9 of **1H'** on the carbocationic intermediate of the $\Delta^8 \rightarrow \Delta^7$-isomerization. In this superimposition, the N atom of **1H** is fitted in with the C-8 of the carbocationic sterol according to the initial hypothesis. The benzyl and 4-(4-chlorophenoxy)benzyl groups are respectively oriented to the ring A and the side-chain at the C-17 of the sterol so that each pair of the corresponding moieties overlaps as far as possible. The superimposition was made manually because it was very difficult to identify reference atoms between two molecules to be compared in this case. The ammonium H atom is located on the α-side of the sterol skeleton, because this side in the vicinity of the C-8 atom was shown to be electrostatically more positive than the β-side by the MO calculation. In Figure 8, the ammonium ion **1H** in the (*S*) configuration and the carbocationic intermediate of the $\Delta^8 \rightarrow \Delta^7$-isomerase reaction are superimposed reasonably well, indicating a similarity in shape between the fecosterol cation and **1H** in its active conformation derived from conformer 9 of **1H'**. The molecular volume occupied in common by the two cations superimposed as shown in Figure 8 was calculated to be 183 Å^3 which was 65(=183/281) and 57(=183/321)% of the molecular volume of **1H** and the fecosterol cation, respectively. A careful superimposition of **1H** in the conformation derived from conformer 10 of **1H'** on the fecosterol cation gave the common volume of 174 Å^3 which was 62 and 54% of the molecular volume of **1H** and the carbocationic intermediate, respectively. This indicates that the 3-D structure of **1H** derived from conformer 10 of **1H'** is slightly less superimposable on the carbocationic intermediate. In each of (*R*)-**3H** and **4H**, one candidate for the active conformation was derived by the same procedure as in the case of **1H** described above, where both of the conformers of **3H'** and **4H'** used

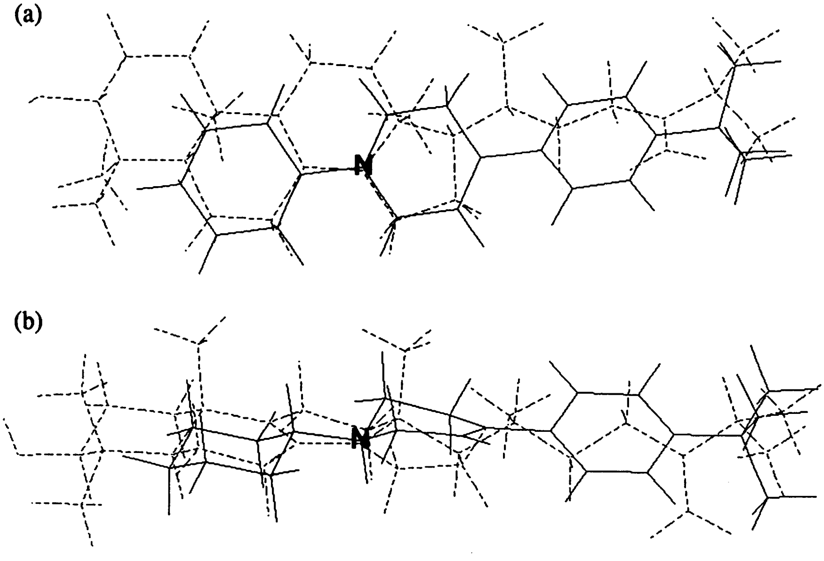


Figure 10. A candidate for the active conformer of the ammonium ion **2H** (solid line) and its superimposition on the carbocationic intermediate (dotted line) of the Δ^{14}-reductase reaction. (a) and (b) are front and side views, respectively.

for the derivation were similar to conformer 9 of (*S*)-**1H'**. Both of the candidate conformations derived were superimposable quite well on the carbocationic intermediate. The common molecular volume occupied by (*R*)-**3H** and the carbocationic intermediate superimposed as carefully as possible was 180 Å^3, and its percentages relative to the molecular volume of the two ions were 64(=180/281) and 56(=180/321)%, respectively. The corresponding values in the superimposition of **4H** and the carbocationic intermediate were 177 Å^3, 66(=177/269)%, and 55(=177/321)%, respectively.

The superimposition of the (*S*) ammonium ion **2H** on the carbocationic intermediate of the Δ^{14}-reductase reaction is displayed in Figure 10. The N atom of the ammonium ion **2H** is fitted in with the sterol C-14 atom, and the cyclohexyl and *para-t*-butylphenyl rings are respectively oriented to the ring B and side-chain of the sterol. The ammonium H atom is located on the α-side of the sterol intermediate as in the case of the ammonium ion **1H**. The MO calculation showed that the α-side was electrostatically much more positive in the neighborhood of the C-14 atom of the carbocationic intermediate. In Figure 10, the ammonium ion **2H** and the carbocationic intermediate of the Δ^{14}-reductase reaction are superimposed satisfactorily well, where the common molecular volume occupied by the two cations is 173 Å^3 and its percentages relative to their molecular volume are 69(=173/250) and 50(=173/344)%, respectively. This also indicates a similarity in shape between the intermediate cation for the Δ^{14}-reduction and **2H** in its adopted active conformation.

Akers *et al.* pointed out the allylic delocalization of the positive charge of the carbocationic intermediate in the Δ^{14}-reductase reaction and examined the correspondence of the N atom of fenpropimorph to the C-14, C-9, and C-8 atoms of the sterol in the 2-D level (*49*). They synthesized analogs of fenpropimorph, which are regarded as 14-, 9-, and 8-azasterol mimics, having inhibitory activity against the Δ^{14}-reductase. We calculated atomic net charges of the carbocationic intermediate from electrostatic potentials on the basis of the algorithm of Chirlian and Francl (*50*) using the SPARTAN system version 3 (1993) (*51*). The calculated charge of the C-14 and C-9 atoms was +0.30 and +0.42, respectively, whereas that of the C-8 atom was –0.38. Therefore, we examined again the superimposition of the ammonium ion **2H** and the carbocationic intermediate of the Δ^{14}-reductase reaction by fitting in the N atom with the sterol C-9 atom. In this case, two molecules were also found to be superimposed reasonably well, giving the common molecular volume of 193 Å^3 which is 77 and 56% of the molecular volume of **2H** and the carbocationic intermediate, respectively. Thus, the N atom of compound **2** also corresponds to the sterol C-9 atom. This N atom is originally placed in the location corresponding to the position 9 of the sterol structure in the initial designing stage. The designing idea was, in a sense, supported by the result of superimposition, although the enzyme originally chosen as the target was the $\Delta^8 \rightarrow \Delta^7$-isomerase but not the Δ^{14}-reductase.

Conclusion

An example of the application of the ACACS molecular modeling system to the fungicide research in the Sumitomo Chemical was presented above. In this example, the new antifungal tertiary amines described may be favorable in safety aspects, since they belong to the same compound series in a wide sense as do fenpropimorph and fenpropidin on the market having low mammalian toxicity (*e.g.*, the acute oral LD_{50} values for rats: *c.* 3000 and 1800 mg/kg (*33*), respectively) and have the same mode of action. The conformational analyses of the new amine compounds were performed after the synthetic work was over. Moreover, the candidate compounds were not developed further. However, we believe that the experience gained during the analyses will be rewarded in designing new agrochemicals having not only the steroidal skeleton but also other types of structures.

Acknowledgments

We greatly appreciate Professor Toshio Fujita of Kyoto University for his invaluable discussions and suggestions concerning computer-aided agrochemical design researches during the last 25 years. We also acknowledge Professor Corwin Hansch of Pomona College for giving us the chance to present our work in this book.

Literature Cited

1. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964,** *86,* 1616.
2. Marsili, M.; Yoshida, M.; Kikuzono, Y.; Katsumi, H. In *Computer Aided Innovation of New Materials;* Doyama, M.; Suzuki, T.; Kihara, J.; Yamamoto, R., Eds.; Elsevier Science Publishers: Amsterdam, 1991; pp 423–428.
3. Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. In *Chemometrics. Mathematics and Statistics in Chemistry;* Kowalski, B. R., Ed.; D. Reidel Publishing: Dordrecht, 1984; pp 17–95.
4. Yoshida, M.; Ueda, A.; Kikuzono, Y.; Takayama, C.; Mizutani, T.; Motoki, T.; Morooka, S.; Shimoju, T.; Kanaoka, S.; Yokota, A. *Abstr. Int. Chem. Cong. Pacific Basin Soc. (Honolulu)* **1984,** 08A23.
5. Kurita, Y.; Takayama, C.; Katagi, T.; Yoshida, M. In *Computer Aided Innovation of New Materials;* Doyama, M.; Suzuki, T.; Kihara, J.; Yamamoto, R., Eds.; Elsevier Science Publishers: Amsterdam, 1991; pp 455–460.

6. Yoshida, M. In *Lead Generation of New Drugs;* Moriguchi, I.; Umeyama, H., Eds.; Tokyo Kagaku Dojin: Tokyo, 1987; pp 37–48.
7. Fujita, T. In *Trends in QSAR and Molecular Modelling 92;* Wermuth, C. G., Ed.; ESCOM Science Publishers: Leiden, 1993; pp 143–159.
8. Molecular Design Ltd., San Leandro, CA.
9. ORACLE Corporation, Redwood Shores, CA.
10. Hansch, C. *Acc. Chem. Res.* **1993,** *26,* 147.
11. Fujinami, A.; Satomi, T.; Mine, A.; Fujita, T. *Pestic. Biochem. Physiol.* **1976,** *6,* 287.
12. Fujinami, A. *Weed Res. (Jpn.)* **1976,** *21,* 153.
13. Kirino, O. *J. Pestic. Sci.* **1984,** *9,* 571.
14. Kamoshita, K.; Kirino, O. In *Structure-Activity Relationships–Quantitative Approach: The Application to Drug Design and Mode-of-Action Studies;* Fujita, T., Ed.; Nankodo: Tokyo, 1982; pp 203–227.
15. Takayama, C.; Kamoshita, K.; Fujinami, A.; Kirino, O. In *Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals;* Draber, W.; Fujita, T., Eds.; CRC Press: Boca Raton, FL, 1992; pp 315–330.
16. Takayama, C. *J. Pestic. Sci.* **1985,** *10,* 591.
17. Takahashi, J.; Kirino, O.; Takayama, C.; Nakamura, S.; Noguchi, H.; Kato, T.; Kamoshita, K. *Pestic. Biochem. Physiol.* **1988,** *30,* 262.
18. Takahashi, J.; Kirino, O.; Takayama, C.; Nakamura, S.; Noguchi, H.; Kato, T.; Kamoshita, K. *J. Pestic. Sci.* **1988,** *13,* 587.
19. Kamoshita, K.; Takayama, C.; Takahashi, J.; Fujinami, A. In *Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals;* Draber, W.; Fujita, T., Eds.; CRC Press: Boca Raton, FL, 1992; pp 429–444.
20. Matsumoto, H. *Jpn. Pestic. Inform.* **1987,** *No. 50,* 20.
21. Moriguchi, I.; Komatsu, K. *Eur. J. Med. Chem.–Chim. Ther.* **1981,** *16,* 19.
22. Kurita, Y.; Tsushima, K.; Takayama, C. In *Probing Bioactive Mechanisms;* Magee, P. S.; Henry, D. R.; Block, J. H., Eds.; American Chemical Society: Washington, D. C., 1989; pp 183–197.
23. Watanabe K.; Umeda, K.; Kurita, Y.; Takayama, C.; Miyakado, M. *Pestic. Biochem. Physiol.* **1990,** *37,* 275.
24. Kato, T. In *Sterol Biosynthesis Inhibitors and Anti-Feeding Compounds;* Haug, G.; Hoffmann, H., Eds; Springer-Verlag: Berlin, 1986; pp 1–24.
25. Baloch, R. I.; Mercer, E. I. *Phytochemistry* **1987,** *26,* 663.
26. Huxley-Tencer, A.; Francotte, E.; Bladocha-Moreau, M. *Pestic. Sci.* **1992,** *34,* 65.
27. Pierce, H. D., Jr.; Pierce, A. M.; Srinivasan, R.; Unrau, A. M.; Oehlschlager, A. C. *Biochim. Biophys. Acta* **1978,** *529,* 429.
28. Avruch, L.; Fischer, S.; Pierce, H., Jr.; Oehlschlager, A. C. *Can. J. Biochem.* **1976,** *54,* 657.
29. Takano, H.; Oguri, Y.; Kato, T. *J. Pestic. Sci.* **1983,** *8,* 575.
30. Baloch, R. I.; Mercer, E. I.; Wiggins, T. E.; Baldwin, B. C. *Phytochemistry* **1984,** *23,* 2219.
31. Rahier, A.; Schmitt, P.; Huss, B.; Benveniste, P.; Pommer, E. H. *Pestic. Biochem. Physiol.* **1986,** *25,* 112.
32. Taton, M.; Benveniste, P.; Rahier, A. *Pestic. Sci.* **1987,** *21,* 269.
33. *The Pesticide Manual;* Worthing, C. R.; Hance, R. J., Eds.; 9th Ed.; British Crop Protection Council: Fahnham, Surrey.
34. Takayama, C.; Fujita, T.; Nakajima, M. *J. Org. Chem.* **1979,** *44,* 2871.
35. Hall, H. K., Jr. *J. Am. Chem. Soc.* **1957,** *79,* 5441.
36. Taft, R. W., Jr. In *Steric Effects in Organic Chemistry;* Newman, M. S., Ed.; John Wiley & Sons: New York, 1956; pp 556–675.

37. Hancock, C. K.; Meyers, E. A.; Yager, B. J. *J. Am. Chem. Soc.* **1961,** *83,* 4211.
38. Takayama, C.; Fujinami, A.; Kirino, O.; Kato, T. *J. Pestic. Sci.* **1983,** *8,* 193.
39. Nakagawa, Y.; Akagi, T.; Iwamura, H.; Fujita, T. *Pestic. Biochem. Physiol.* **1988,** *30,* 67.
40. Leffler, J. E.; Grunwald, E. *Rates and Equilibria of Organic Reactions;* John Wiley & Sons: New York, 1963; pp 171–262.
41. Stewart, J. J. P. *J. Comput. Chem.* **1989,** *10,* 221.
42. Stewart, J. J. P.; Fujitsu Ltd., Tokyo.
43. Carlisle, C. H.; Timmins, P. A. *J. Cryst. Mol. Struct.* **1974,** *4,* 31.
44. Jensen, J. S.; Jørgensen, F. S.; Klemmensen, P. D.; Hacksell, U.; Pettersson, I. *Pestic. Sci.* **1992,** *36,* 309.
45. Lipton, M.; Still, W. C. *J. Comput. Chem.* **1988,** *9,* 343.
46. Kabsch, W. *Acta Cryst.* **1976,** *A32,* 922.
47. Leo, A. J. *Chem. Rev.* **1993,** *93,* 1281.
48. Grdadolnik, S. G.; Mierke, D. F.; Byk, G.; Zeltser, I.; Gilon, C.; Kessler, H. *J. Med. Chem.* **1994,** *37,* 2145.
49. Akers, A.; Ammermann, E.; Buschmann, E.; Götz, N.; Himmele, W.; Lorenz, G.; Pommer, E.-H.; Rentzea, C.; Röhl, F.; Siegel, H.; Zipperer, B.; Sauter, H.; Zipplies, M. *Pestic. Sci.* **1991,** *31,* 521.
50. Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987,** *8,* 894.
51. Wavefunction, Inc., Irvine, CA.

RECEIVED April 6, 1995

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Chapter 13

QSARs and Three-Dimensional Shape Studies of Fungicidal Azolylmethylcyclopentanols

Molecular Design of Novel Fungicides Metconazole and Ipconazole

Hiroshi Chuman[1], Atsushi Ito[2], Toshihide Saishoji[2], and Satoru Kumazawa[2]

[1]**Intelligent Systems Development for Research Department, Kureha Chemical Industry Company, Ltd., 3–25–1 Hyakunin-cho, Shinjuku-ku, Tokyo 169, Japan**
[2]**Nishiki Research Laboratory, Kureha Chemical Industry Company, Ltd., 16 Ochiai, Nishiki-machi, Iwaki-shi, Fukushima 974, Japan**

A series of azole compounds containing a cyclopentane ring were synthesized and tested for fungicidal activity. The Hansch-Fujita type QSAR and three-dimensional shape comparison analyses were employed to optimize the structure of a lead compound logically for the higher fungicidal activity, resulting in the discovery of metconazole and ipconazole, promising fungicides of novel structure. A possible mode of interaction between metconazole and its target receptor, cytochrome $P450_{14DM}$ was proposed.

Since the end of 1960s, a number of so-called azole type fungicides, which show their antifungal activity by inhibiting the C-14 demethylation step of the ergosterol biosynthetic pathway, have been investigated in both academia and industry over the world (*1*). Hundreds of patents covering azole type fungicides have been published and over twenty new compounds have been introduced commercially for agricultural use. As seen in the structures of a number of azole type fungicides, the most general structure requirements for the activity is considered to contain a nitrogen-containing heteroaromatic (azole) ring, a hydroxy group and a benzene ring. According to the number of chemical bonds intervening between the azole and the benzene rings, fenarimol (I) and fluotrimazole (II) are classified in the two-bond type, and most of other typical fungicides such as triadimefon (III), propiconazole (IV), cyproconazole (V), diniconazole (VI), flusilazole (VII) and triflumizole (VIII) are classified in the three-bond type, as shown in Figure 1.

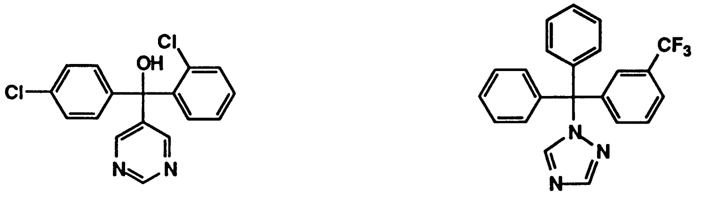
We have synthesized a series of azolylmethylcyclopentanols, classified as the five-bond type and measured their antifungal activities against plant pathogens *in vitro* and *in vivo*. Among them, we have discovered metconazole (IX) and ipconazole (X) also shown in Figure 1, which show a high fungicidal activity and a broad spectrum against various plant diseases. The safety studies have confirmed their low-toxicity for mammals, birds, and aquatic organisms and also have proved their degradability in the environment. Metconazole (IX) was launched in the spring of 1994 in France for

0097–6156/95/0606–0171$12.00/0

controlling important cereal pests such as septoria and fusarium diseases, and rusts and leaf blotches. At the same time, ipconazole (X) was introduced for controlling seedborne rice diseases in Japan.

This paper focuses on the lead optimization processes, during which the traditional Hansch-Fujita type analyses of the quantitative structure-activity relationships (QSAR) and three-dimensional shape comparisons gave essential contributions to achieving the discovery.

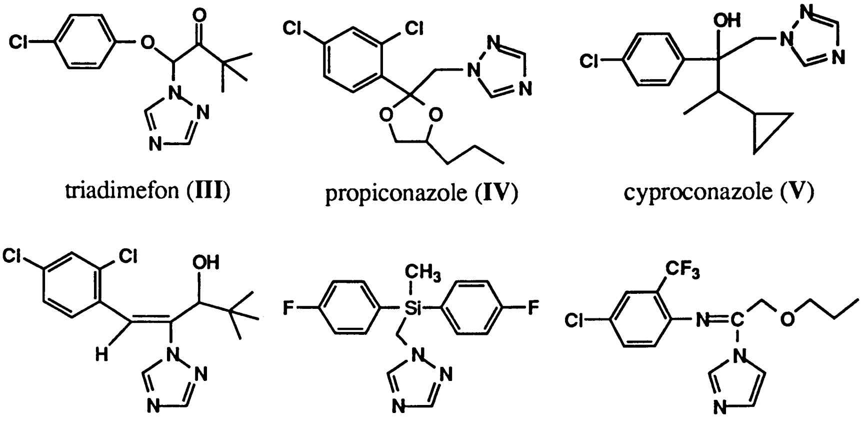
Two-bond type :



fenarimol (I)

fluotrimazole (II)

Three-bond type :



triadimefon (III)

propiconazole (IV)

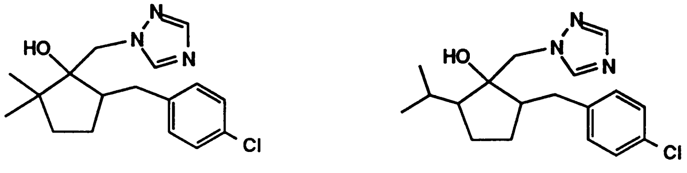
cyproconazole (V)

diniconazole (VI)

flusilazole (VII)

triflumizole (VIII)

Five-bond type :



metconazole (IX)

ipconazole (X)

Figure 1. Typical and novel azole type fungicides for agricultural use.

Materials and Methods

Compounds : The synthesis of compounds **1–34** is briefly outlined in Figure 2 (for details, see Jpn. Kokai Tokkyo Koho JP62-149667, 1987; JP1-93574, 1989). The compounds were classified into types I and II according to those without and with the R_1, R_2-substituents, respectively, and listed in Tables I and II separately. Each diastereo-isomeric set containing enantiomers (racemate) was isolated at the step of the epoxide intermediate. The configuration of the triazolyl compounds shown in Figure 2 is that of the active form. Compounds **1–32** were used for the QSAR studies. Compounds **19, 33** and **34** were optically resolved by the normal phase high-performance liquid chromatography (HPLC) with a chiral column (CHIRALCELL OD 500 mm x 10 mm i.d., DAICEL Chemical Industries Ltd.). The chemical purity (p), specific rotation and enatiometric excess (ee) were as follows : (–)-**19** ; p = 100.0 %, $[\alpha]^{D}_{20}$ = –23.7° (c = 10.0, EtOH), ee = 100.0 %, (–)-**33** ; p = 99.9 %, $[\alpha]^{D}_{20}$ = –3.3° (c= 1.23, EtOH), ee = 99.8 %, (–)-compound **34** ; p = 98.2 %, $[\alpha]^{D}_{20}$ = –24.4° (c = 1.11, EtOH), ee = 96.4 %. (±)-compounds **19** and **33** are, respectively, metconazole and ipconazole.

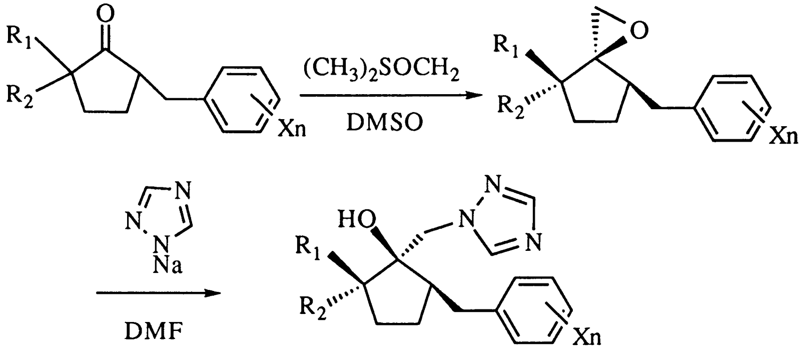
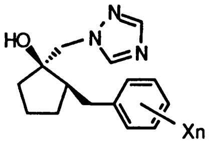


Figure 2. Synthetic scheme. For absolute configuration, see below.

Antifungal Activities *in vitro* : Each test compound was dissolved and diluted with dimethyl sulfoxide (DMSO), and then suspended in sterilized potato sucrose agar (PSA) medium to give prescribed concentrations (the final DMSO concentration was 1 %). The medium was poured in petri dishes and was hardened. Mycelial discs (4 mm in diameter) of *Botrytis cinerea* and *Gibberella fujikuroi* grown on the PSA alone in advance were transferred to the PSA plates containing the various concentrations of each compound and incubated. The diameter of each colony was measured to determine the IC_{50} value (the effective concentration in M for 50 % inhibition of mycelial growth). The pIC_{50} (*in vitro*) is listed in Tables I and II. For some compounds, the minimum inhibitory concentration (MIC) in μg/ml against various fungal species was recorded. The incubation was performed at 20 or 28°C for 2 or 3 days, depending on fungal species.

Antifungal Activities *in vivo* : The each type I compound was dissolved in acetone and suspended in water (the final acetone concentration was 5 %) with 60 μg/ml of a detergent, Gramin S. The suspension was diluted with water containing 5% of acetone and 60μg/ml of Gramin S to give a series of prescribed concentrations and was sprayed on a cucumber seedling at the first leaf stage. After 4 hours of the chemical treatment, two mycelial discs (4 mm in diameter) of *Botrytis cinerea* grown on the PSA

Table I. Antifungal Activities and Physicochemical and Steric Parameters of Triazolylmethylcyclopentanols (Type I)

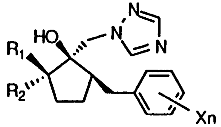


Compounds[a]		Log P[b]	Steric Parameters of PhXn (Å)[c]				pIC_{50}[d] (*in vitro*)						(*in vivo*)	
			L1	L2	W	D	*B.c.*[e]			*G.f.*[f]			*B.c.*[e]	
No.	Xn						obsd.	calcd.[g]	calcd.[h]	obsd.	calcd.[i]	calcd.[j]	obsd.	calcd.[k]
1	H	2.26*	3.68	2.48	6.70	3.48	4.6	5.01	4.97	4.8	5.12	5.08	2.9	3.24
2	2-Cl	2.95	3.68	3.30	7.79	3.50	3.8	3.83	4.12	4.3	4.37	4.83	1.9	2.41
3	3-Cl	2.98	3.68	2.48	7.79	3.50	4.7	4.93	4.77	5.0	5.08	4.85	1.9	2.41
4	3-CF_3	3.07	4.11	2.48	8.29	4.83	3.9	3.62	3.51	4.2	4.21	4.13	1.9	1.83
5	4-Cl	3.11*	4.86	2.48	6.70	3.50	6.1	5.95	5.80	5.9	5.95	5.77	3.8	3.55
6	4-F	2.39	4.22	2.48	6.70	3.48	5.1	5.21	5.14	5.3	5.28	5.23	3.3	3.38
7	4-Br	3.27*	5.10	2.48	6.70	3.70	5.9	5.89	5.75	5.9	5.96	5.77	4.1	3.45
8	4-CF_3	3.07	4.85	2.48	6.70	4.83	5.3	4.97	4.88	—[l]			3.2	3.57
9	4-CH_3	2.71*	4.68	2.48	6.70	3.97	5.4	5.26	5.16	5.5	5.42	5.40	3.5	3.58
10	4-*t*-Bu	3.97*	5.75	2.48	6.71	6.16	3.9	4.20	4.25	5.1	5.11	5.06	1.9	2.37
11	4-Ph	3.92*	7.96	2.48	6.70	3.48	6.2	6.13	6.05	—[l]			2.2	2.49
12	4-CN	1.70	5.56	2.48	6.70	3.48	4.2	3.96	4.09	4.3	4.29	4.25	2.5	2.24
13	2,4-Cl_2	3.87	4.86	3.30	7.79	3.50	4.4	4.13	4.47	4.9	4.80	4.96	1.9	1.40
14	2,4-F_2	2.60	4.22	2.84	7.20	3.48	4.7	4.59	4.67	5.0	4.86	5.04	3.2	2.99
15	2-F,4-Cl	3.27	4.86	2.84	7.20	3.50	5.2	5.14	5.18	5.5	5.39	5.43	3.3	2.90
16	2,6-F_2	2.60	3.68	2.84	7.70	3.48	3.7	4.17	4.24	4.3	4.51	4.65	2.1	2.45
17	3,4-F_2	2.51	4.22	2.48	7.20	3.48	5.3	4.95	4.85	5.6	5.07	4.96	3.3	2.93
18	2,3,4,5,6-F_5	3.16	4.22	2.84	7.70	3.48	4.5	4.67	4.70	4.8	4.98	5.01	2.7	2.43

For footnotes, see Table II.

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Table II. Antifungal Activities and Physicochemical and Steric Parameters of Triazolylmethylcyclopentanols (Type II)



Compounds [a]				Log P [b]	Steric Parameters of PhXn (Å) [c]				pIC_{50} [d] (*in vitro*)			
					L1	L2	W	D	*B.c.* [e]		*G.f.* [f]	
No.	R1	R2	Xn						obsd.	calcd.[h]	obsd.	calcd.[j]
19	Me	Me	4-Cl	3.83*	4.86	2.48	6.70	3.50	6.6	6.47	6.8	6.84
20	Me	Me	4-F	3.11	4.22	2.48	6.70	3.48	6.4	6.24	6.6	6.79
21	Me	Me	4-Br	3.99	5.10	2.48	6.70	3.70	6.5	6.33	6.9	6.74
22	Me	Me	4-CH_3	3.43	4.68	2.48	6.70	3.97	6.5	6.07	6.8	6.75
23	Me	Me	4-*t*-Bu	4.69	5.75	2.48	6.71	6.16	3.8	4.40	5.5	5.55
24	Me	Me	4-Ph	4.64	7.96	2.48	6.70	3.48	6.1	6.23	6.1	6.33
25	Me	Me	2-F,4-Cl	3.99	4.86	2.84	7.20	3.50	6.2	5.75	6.7	6.39
26	Me	Me	3-OH,4-Cl	3.16	4.86	2.48	7.27	3.50	5.1	5.76	6.7	6.35
27	Et	H	4-Cl	4.02*	4.86	2.48	6.70	3.50	6.4	6.46	6.7	6.78
28	Et	H	4-F	3.30	4.22	2.48	6.70	3.48	6.0	6.34	6.6	6.85
29	Et	H	4-Br	4.18	5.10	2.48	6.70	3.70	6.4	6.29	6.7	6.64
30	Et	H	2,4-Cl_2	4.78	4.86	3.30	7.79	3.50	5.1	4.55	5.5	5.31
31	Et	H	4-*t*-Bu	4.88	5.75	2.48	6.71	6.16	4.6	4.25	5.3	5.33
32	Et	H	4-Ph	4.83	7.96	2.48	6.70	3.48	5.9	6.51	6.2	6.12
33	*iso*-Pr	H	4-Cl	4.28*					6.4		6.9	
34	H	Et	4-Cl	3.71*					—[l]		—[l]	

a) Tested as the racemate of enatiomers the configuration of which is shown above. b) Asterisked values were measured.
c) Definitions of steric parameters are described in the text. d) $pIC_{50} = -\text{Log } IC_{50}$ (M)
e) *B.c.* : *Botrytis cinerea* f) *G.f.* : *Gibberella fujikuroi* g) Calculated by equation 4. h) Calculated by equation 7.
i) Calculated by equation 5. j) Calculated by equation 8. k) Calculated by equation 6. l) — : not measured.

medium alone in advance were inoculated on the first foliage leaf of the cucumber treated with each concentration of the test compounds. The plants were kept in 100 % humidity at 20°C for 3 days. The IC_{50} value (the effective concentration in M in the sprayed suspension for 50 % inhibition of the disease expansion) was determined from the diameter of each infected area. The pIC_{50} (*in vivo*) is listed in Table I.

Partition Coefficient (Log P) : Partition coefficients (P) of type I compounds, **1, 5, 7, 9, 10** and **11,** were measured in the octanol-water system by the shaking-flask method. By using the reversed phase HPLC, k' value for the same set of compounds was also measured . k' is defined as the ratio of retention times : $(t - t_0) / t_0$, where t_0 is the retention time of a reference compound, fructose, and t is that of each test compound. A good linear relationship between log P and log k' was obtained for these compounds as shown in equation 1.

$$\log P = 3.77 \log k' + 2.93 \qquad (1)$$
$$n = 6,\ r = 0.975,\ s = 0.166$$

In this and the following equations 4 – 8, n is the number of data used in the correlation, r is the multiple correlation coefficient and s is the standard deviation. Each term in equations 1 and 4 – 8 was confirmed as being significant above the 95 % level. Log P values of the other type I compounds was calculated from the k' value and equation 1. Log P value of type II compounds, **19, 27, 33** and **34,** were measured by the shaking-flask method. Log P value of the other type II compounds was estimated by the additive rule: *i.e.*, by adding π values of alkyl substituents calculated from measured log P values of compounds **5, 19** and **27** to the log P of the compounds without alkyl substituents. Tables I and II list log P value of compounds used in this study. The log P value of typical azole fungicides, **III, IV, VI, VII** and **VIII,** were newly measured as shown below.

Steric Parameters : In order to describe steric influences of substituents in the phenyl moiety, we introduced steric parameters similar to the STERIMOL (2) in our QSAR equations. Definitions of steric parameters are shown in Figure 3. L_1 and L_2 denote the maximum length of *para*- and *ortho*-substituents from the center of the benzene ring along the x-axis direction, respectively. W and D are the measure of the maximum width of the substituted benzene ring in two directions perpendicular to the x-axis. The relevant steric parameters are listed in Tables I and II.

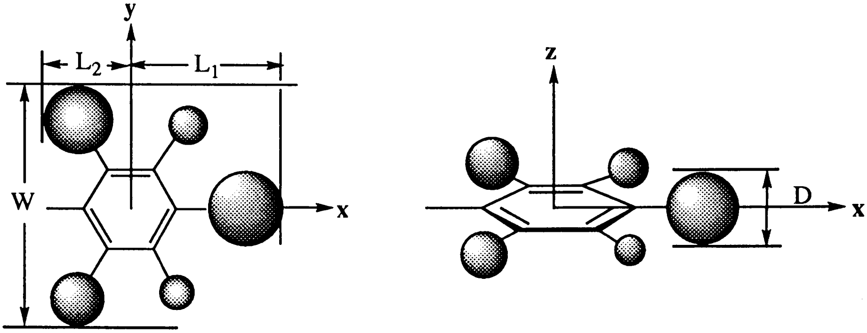


Figure 3. Definition of the steric parameters.

Superposition and Conformation Analyses : The least-squares superposing in terms of corresponding atomic positions is the most widely used method. This method superposes two molecules, A and B, by minimizing a target function F* indicated as equation 2.

$$F^* = \Sigma\, \omega_i \mid r_{Ai} - r_{Bi} \mid^2 \qquad (2)$$

r_{Ai} (r_{Bi}) : *i* th corresponding atomic position vector of the molecule A (B).
ω_i : weight of *i* th corresponding pair.

It should be noted, however, that this method is generally limited to rigid molecular systems. The function value of F* is sensitive to small variations in conformation. It is generally difficult to find the best superposition because the value of F* depends on a given conformation of the molecule A (B). The compounds treated here contain rotatable bonds around which conformational changes affect the value of F*. In order to consider this conformational flexibility, we introduced torsion angle variables that express conformational freedom of molecules and we also added a term arising from the conformational energy change, $\Delta E(\Theta_A)$ ($\Delta E(\Theta_B)$) of the molecule A (B) (Chuman, H., unpublished work).

$$F = \Sigma\, \omega_i \mid r_{Ai}(\Theta_A) - r_{Bi}(\Theta_B) \mid^2 + S\, [\, \Delta E(\Theta_A) + \Delta E(\Theta_B)\,] \qquad (3)$$

$r_{Ai}(\Theta_A)$ [$r_{Bi}(\Theta_B)$] : *i* th corresponding atomic position vector, r_{Ai} (r_{Bi}), is a function of torsion angles, Θ_A (Θ_B) for the molecule A (B).
S : scaling factor in the dimension of square distance / energy.

For minimizing the target function F, the first term only with fixed torsion angles was first minimized by the Monte Carlo method which overcomes local minimum problems (*4*) and then the minimization of the total F value as the function of the torsion angles was performed. The second term in equation 3 was estimated by a force field similar to ECEPP (*3*) in which independent variables are torsion angles only.

In computational conformation analysis, we generated initial conformers by a combination of the grid search (*4*) (systematically changing the angular increment of each rotatable bond) and the ring conformation search (*5*) (a systematic conformation search with ring closure constraints), and then optimized these structures with MNDO (*6*). In the superposition of molecules mentioned above, each optimized structure obtained by MNDO was used as the starting conformation in which torsion angles were initial values of the independent variable set, Θ, in the target function F.

Results and Discussion

QSAR(I) : Series of type I compounds (**1–18**) with various substituents in the phenyl moiety were assayed for the antifungal activity against *Botrytis cinerea in vitro* as well as *in vivo* and against *Gibberella fujikuroi in vitro*. The results seem to show that the optimal substituents on the phenyl moiety are *para*-chloro and bromo. The corresponding compounds (**5** and **7**) are among the most active *in vitro* as well as *in vivo* as seen from Table I. To validate this, variations in the antifungal activity were examined using physicochemical and steric parameters and regression analyses. Electronic parameters of Xn substituents such as the Hammett σ were not statistically significant. Equations 4–6 gave the best correlation quality among various combinations of log P, the steric parameters, L_1, L_2, W and D, and squared terms of these parameters. Negative signs of parameter terms in equations 4–6 suggest that the phenyl moiety is accommodated inside a sterically restricted hydrophobic pocket of the target receptor site. The difference in the optimum log P value for the activity against *Botrytis cinerea* between *in vitro* and *in vivo* (3.71 and 2.87, respectively) seems to reflect a difference in the transport process to the target site under respective conditions. While the favorable substitution patterns in some other azole type fungicides are 2,4-dihalo (7) as found in propiconazole (**IV**), diniconazole (**VI**) and triflumizole (**VIII**), the QSAR results show the substituent on the 2 (*ortho*) position of the benzene ring is definitely unfavorable for the activity because of the negative coefficients of the L_2 term in equations 4 and 5. The introduction of bulkier substituents to the *para* position is also unfavorable because of the negative W and D terms. Only introduction of a moderately hydrophobic substituent into the *para* position seems to enhance the activity because of its dependence on the log P. With these QSAR considerations, we confirmed and selected the *para*-chloro derivative (compound **5**) as the most promising lead compound for the next step.

pIC_{50} (*Botrytis cinerea in vitro*)

$$= -1.309\,L_2 - 0.843\,W - 0.716\,D + 4.023\log P - 0.541(\log P)^2 + 10.068 \quad (4)$$

$n = 18,\ r = 0.952,\ s = 0.282,\ \log P^{opt} = 3.71$

pIC_{50} (*Gibberella fujikuroi in vitro*)

$$= -0.833\,L_2 - 0.711\,W - 0.435\,D + 2.896\log P - 0.356(\log P)^2 + 8.735 \quad (5)$$

$n = 16,\ r = 0.938,\ s = 0.237,\ \log P^{opt} = 4.07$

pIC_{50} (*Botrytis cinerea in vivo*)

$$= -1.090\,W + 5.781\log P - 1.008(\log P)^2 + 2.624 \quad (6)$$

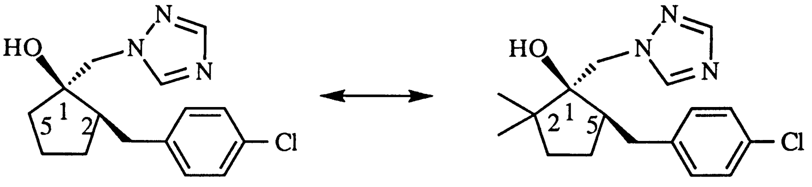
$n = 18,\ r = 0.854,\ s = 0.425,\ \log P^{opt} = 2.87$

The log P value of the lead compound **5**, 3.11, is approximately 0.8 smaller than the averaged value (3.88) of typical fungicides **III** (log P = 3.11), **IV** (3.54), **VI** (4.30), **VII** (3.97) and **VIII** (4.50) in Figure 1, and the calculated molecular volume of compound **5**, 262 Å^3, is also slightly smaller than the averaged volume (*ca.* 280 Å^3) of the same set of compounds. Considering these physicochemical structural features, we moved to the next logical step for the further enhancement of the activity.

Three-Dimensional (3D) Shape Comparison : Compound **5** has two asymmetric carbons and hence two racemic geometrical isomers. The relative configuration of the isomer responsible for the antifungal activity was examined by ^{1}H NMR-NOE experiments which measure NOE signals between benzyl and triazol-1-ylmethylene protons in two geometrical isomers. No intensity increase according to the NOE effect was observed in the more active isomer, whereas there was a considerable increase in the less active one. The relative configuration of the active isomer of compound **5** was thus concluded to be (1*S*,2*S*) or (1*R*,2*R*) (the *cis* arrangement of the hydroxy and benzyl groups as shown in Figure 2 and Table I).

We planned to increase the hydrophobicity and molecular volume of compound **5** for higher antifungal activity. A number of such analogs are conceivable. However, the results of the QSAR(I) indicate no more space for introducing hydrophobic substituents into the phenyl moiety. Therefore, as a possible choice, we tried to introduce ***gem***-dimethyl at the 5 position of the cyclopentane ring of compound **5** leading to compound **19** (metconazole) (note changes in the stereochemical notation and the atom numbering of compounds **5** and **19**, see Figure 4).

In spite of the broad diversity in the 2D structure of azole type fungicides, their 3D structures are assumed to take similar shapes which are complementary to the binding site in the target receptor. From this assumption, we carried out the 3D shape comparison calculation between compound **19** and diniconazole (**VI**) as the reference



(1*R*,2*R*)-compound **5**

(1*S*,5*R*)-compound **19**

Figure 4. The stereochemical and the atom numbering relation between compounds **5** and **19** (metconazole).

simultaneously with the synthetic effort of compound **19**. Diniconazole is reported(*8*,*9*) to show a high antifungal activitiy (pIC_{50} (M) *in vitro* for *Botrytis cinerea* = 5.5 and for *Gibberella fujikuroi* = 5.9 under slightly different experimental conditions). Each energy minimum conformer of compound **19** obtained by MNDO was compared to that of diniconazole by the superposition method described above. The active configuration of diniconazole was reported as (*R*) (common name : diniconazole M) as shown in the right hand side of Figure 5. Katagi (*10*) reported the existence of two major conformers, I and II, which are different in the torsion angle around the C-C bond to which *tert*-butyl and hydroxy groups are attached geminally. Without a large change in the conformational energy (the second term in equation 3) from that of the starting conformer (local energy minimum conformer obtained by MNDO), the (1*S*,5*R*) configuration of compound **19** was superposable on the conformer II of (*R*)-diniconazole with the smallest root mean square value and the second largest common volume among the four possible pairs as shown in Table III. Figure 5 shows the *gem*-dimethyl of compound **19** and *tert*-butyl group of diniconazole appear to occupy almost the same geometric region. Thus, we considered that the alkyl groups at the 2 position on the cyclopentane ring of compound **19** play an important role for the antifungal activity. Simultaneously, the active configuration of compound **19** was assumed as (1*S*,5*R*) and its active conformation is shown in Figure 5.

Table III. Three-dimensional Shape Comparison between Isomers of Compound 19 (Metconazole) and Those of Diniconazole [a]

Compound **19** [b]	Diniconazole [b]	RMSD [c] (Å)	Common volume ($Å^3$)
(1*S*,5*R*)	(*R*) conformer I	1.13	203
(1*S*,5*R*)	(*R*) conformer II	0.57	211
(1*R*,5*S*)	(*R*) conformer I	0.75	220
(1*R*,5*S*)	(*R*) conformer II	0.87	196

a) 8 pairs of corresponding atoms circled in structural formulae in Figure 5 were used in this comparison.

b) As those in which the value of F in equation 3 is minimum.

c) Root Mean Square Deviation values.

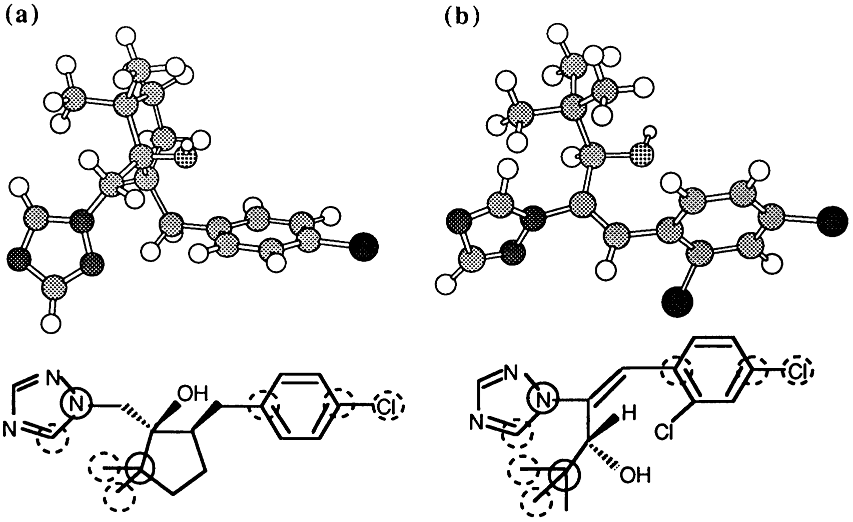


Figure 5. Three-dimensional shape comparison of the (1*S*,5*R*)-isomer of compound **19** (metconazole) (a) and the conformer II of (*R*)-diniconazole (b). Atoms enclosed by circles are used in the three-dimensional superposition and the value ω_i in equation 3 is assigned as two for the atoms in solid and unity for ones in broken circles, respectively.

The optical isomers of compound **19** were prepared and their *in vitro* antifungal activity was measured against a number of phytopathogenic fungi. Table IV lists the MIC values in comparison with that of the lead compound **5**. The (–) isomer of compound **19** indeed showed an excellent activity with a broad fungitoxic spectrum.

Table IV. MIC Values (μg/ml) of Compounds 5 and 19 (Metconazole)

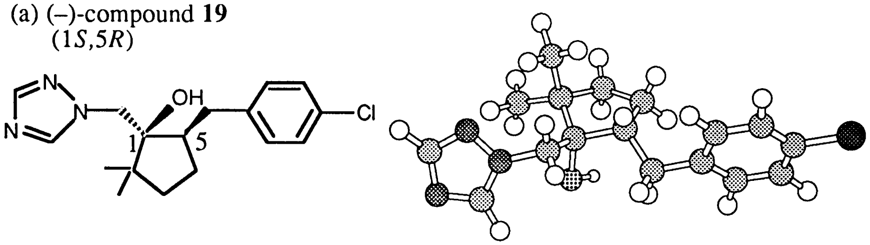
Compounds	Pathogenic Fungi								
	P.o.	*G.f.*	*F.c.*	*B.c.*	*S.s.*	*S.c.*	*G.c.*	*C.b.*	*A.m.*
(±)–**5**	25	6.25	>100	6.25	6.25	0.78	25	12.5	100
(±)–**19**	6.25	0.39	1.56	1.56	1.56	<0.2	1.56	25	25
(–)–**19**	3.13	<0.2	0.78	0.39	0.78	<0.2	0.78	25	12.5
(+)–**19**	100	50	>100	25	>100	1.56	50	>100	>100

P.o. : *Pyricularia oryzae*
F.c. : *Fusarium oxysporum f. sp.cucumerinum*
S.s. : *Sclerotinia sclerotiorum*
G.c. : *Glomerella cingulata*
A.m. : *Alternaria mali*
G.f. : *Gibberella fujikuroi*
B.c. : *Botrytis cinerea*
S.c. : *Sclerotinia cinerea*
C.b. : *Cercospora beticola*

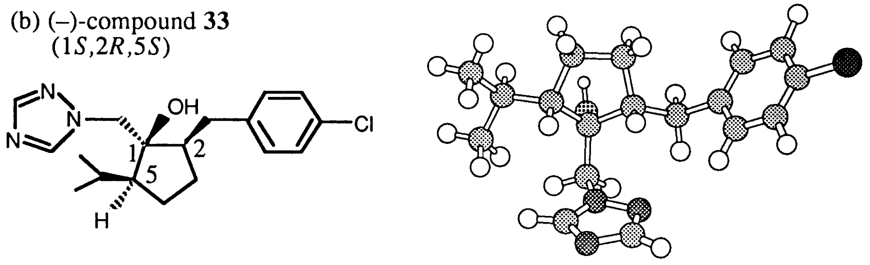
In other words, the introduction of alkyl groups at the 5- position on the cyclopentane ring of the lead compound **5** caused a remarkable increase of the antifungal activity. Our prediction based upon the QSAR and the 3D shape comparison studies was thus confirmed. As another choice to increase the hydrophobicity and substructual molecular volume of compound **5** in its 5-position substituents, we synthesized compound **33** (the isopropyl derivatives, ipconazole) and confirmed its high fungicidal activity. The absolute configuration and active conformation of the antifungally active compounds **19** and **33**, remains to be verified.

X-Ray Crystallographic and Conformational Analyses : In order to assign directly the absolute configuration of antifungally active isomers of the type II compounds, we carried out a single crystal X-ray analysis of the antifungally more active enantiomers of compounds **19**, **33** and **34** (the Et derivative), all of which resolved isomers were optically minus. Their crystal structures were determined as follows : (–)-compound **19** ; monoclinic, $P2_1$, a = 9.015, b = 12.581, c = 7.321 Å, $\gamma = 96.84°$, Z = 2, R = 0.0576 , (–)-compound **33** ; monoclinic, $P2_1$, a = 23.592, b = 8.448, c = 9.973 Å, $\beta = 113.54°$, Z = 4, R = 0.0450, and (–)-compound **34** ; orthorhombic, $P2_12_12_1$, a = 18.132, b = 10.455, c = 8.745 Å, Z = 4, R = 0.0628.

(a) (–)-compound **19**
(1*S*,5*R*)



(b) (–)-compound 33
(1*S*,2*R*,5*S*)



(c) (–)-compound 34
(1*R*,2*R*,5*S*)

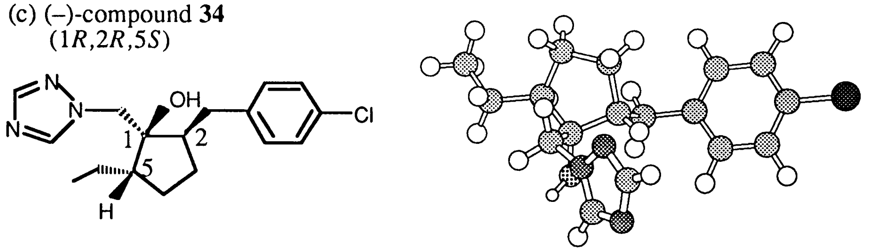


Figure 6. The absolute configuration and the crystal structure of the antifungal (–)-isomers of compounds **19** (metconazole) (a), **33** (ipconazole) (b) and **34** (c).

All the above packing structures contain an intermolecular hydrogen bond between the hydroxyl group and the nitrogen atom at the 4 position of the triazole ring. The X-ray structure analysis assigned the absolute configurations of the above (–)-compounds **19**, **33** and **34** as (1*S*,5*R*), (1*S*,2*R*,5*S*) and (1*R*,2*R*,5*S*), respectively, as shown in Figure 6. These assignments are to verify the above assumption from the 3D shape comparison study. The configuration of four ligands around each of the 1- and 2-carbon atoms in type I compounds and each of the 1-, 2-, and 5-carbon atoms in compounds **19**–**32** is reasonably considered to be the same as that around the corresponding carbon atoms in the above compounds as far as the most active diastereomer is concerned. Therefore, the absolute configuration of compounds as presumed in Figure 2 and Tables I and II seems to be valid.

In compounds listed in Tables I and II, torsion angles, θ_1 and θ_2, of which definitions are shown in Figure 7(a), determine the spatial placement of the two important aromatic rings, triazole and benzene. The values, (θ_1,θ_2) in the crystal structures are (181°, 176°), (52°, 186°) and (–51°, 161°) for (–)-compounds **19**, **33** and **34**, respectively. These angles possibly indicate the general conformational feature of all the compounds discussed in this paper. The conformation with respect to θ_1 for the triazol-1-ylmethyl is considered to be flexible, whereas that with respect to θ_2 for the benzyl is considered to be fixed in the *trans* vicinity to the triazole ring. To investigate the conformation profile and to examine the relevancy of the active conformation predicted by the 3D shape comparison study as shown in Figure 5, we carried out an extensive conformational search of (–)-(1*S*,5*R*)-compound **19** using a semiempirical molecular orbital method, MNDO. We found that there are three major local enery minimum conformers. For these conformers, (θ_1,θ_2) were (191°, 196°), (53°, 187°) and (–66°, 185°). These three sets of (θ_1,θ_2) are close to the observed ones of (–)-compounds **19**, **33** and **34** in the crystal structures, respectively. Energies of these three conformers exist in a small range of values, being 0.8 (181°, 176°), 3.9 (52°, 186°) and 5.4 (–51°, 161°) kcal/mol higher than that of the ground energy minimum, (191°, 196°), respectively. The active conformation predicted in Figure 5(a) takes a folded structure, where (θ_1,θ_2) is located at a region, (*ca.* 60°, *ca.* 60°). This conformation gives a rather high energy value, *ca.*10 kcal/mol from the ground minimum energy. This energy value locates just on the border line for the criterion to select the active conformation as the reasonable one, when errors inherent in MNDO and the neglect of other possible factors affecting the stability such as an energy gain from the binding with the target receptor are taken into account (*11*). In order to clarify this point, we synthesized a conformationally fixed analog, compound **35**, shown in Figure 7(b). Compound **35** takes only the *trans* conformation around the double bond,

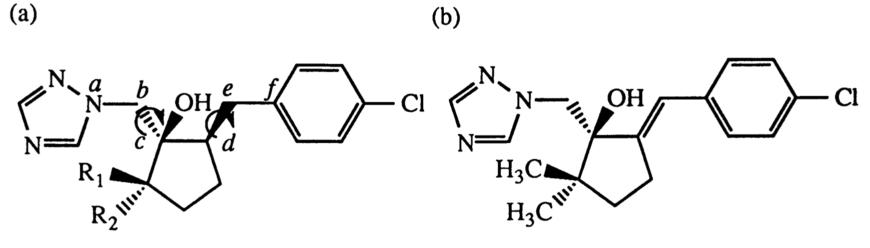


Figure 7. (a) Definition of torsion angles and (b) structure of compound **35**. The torsion angles, θ_1 and θ_2, are defined by the sequential atoms, a-b-c-d and c-d-e-f, respectively.

i.e., $\theta_2 = ca.\ 180°$. Its antifungal activity was much lower (MIC value for *Gibberella fujikuroi* = 6.25 µg/ml) than that of compound **19** (0.39 µg/ml). In general, it is difficult to determine the active conformation explicitly unless a direct X-ray crystallographic analysis of an inhibitor-receptor complex could be performed. However, we tentatively could conclude that the active conformation takes (*ca.* 60°, *ca.* 60°) as the set of torsion angle values, (θ_1,θ_2).

QSAR(II) : The activity potentiating effect of the alkyl substituents on the cyclopentane ring was quantitatively examined to formulate equations 7 and 8, by introducing an indicator variable *I* into equations 4 and 5. The indicator variable, *I* takes zero and unity for type I and type II compounds, respectively. The sign and size of the respective terms, except for the indicator variable in equations 7 and 8, are almost equivalent to those of the corresponding terms in equations 4 and 5. The indicator variable term is statistically significant and the sign is positive. From a point of view independently from the 3D shape comparison results, equations 7 and 8 also support that the alkyl groups on the cyclopentane ring play an important role in binding with the target receptor.

$$\begin{aligned} &\mathrm{pIC}_{50}\ (\textit{Botrytis cinerea in vitro}, \text{type I + type II}) \\ &= -0.761\ L_2 - 0.862\ W - 0.669\ D + 3.237 \log P - 0.418 (\log P)^2 + 0.428\ I \\ &\quad +9.782 \qquad (7) \\ &\quad n = 32,\ r = 0.933,\ s = 0.377,\ \log P^{\,\mathrm{opt}} = 3.87 \end{aligned}$$

$$\begin{aligned} &\mathrm{pIC}_{50}\ (\textit{Gibberella fujikuroi in vitro}, \text{type I + type II}) \\ &= -0.790\ W - 0.270\ D + 3.350 \log P - 0.471 (\log P)^2 + 1.012\ I + 6.145 \qquad (8) \\ &\quad n = 30,\ r = 0.966,\ s = 0.245,\ \log P^{\,\mathrm{opt}} = 3.56 \end{aligned}$$

The binding mode of metconazole and ipconazole to the cytochrome P450$_{14DM}$: Based upon the results obtained, the proposed binding mode of (–)-compound **19**, *i.e.*, (1*S*,5*R*)-metconazole to the target receptor, cytochrome P450 catalyzing lanosterol 14α-demethylation (P450$_{14DM}$) (*12*) is illustrated in Figure 8. The following interactions at the molecular level are proposed :

1. The nitrogen atom at the 4 position of the triazole ring coordinates to the heme ferric ion in P450$_{14DM}$.
2. The presence of the hydroxy group at the 1 position of the cyclopentane ring was experimentally confirmed to be indispensable for the activity (unpublished data). This group is considered to be involved in either hydrogen bonding or electrostatic interaction with an amino acid residue in the close proximity of the active site of P450$_{14DM}$.
3. The QSAR(I) results suggest the existence of a hydrophobic pocket or groove, which accommodates the *para*-chloro-phenyl group, in the proximity of the active site.
4. The 3D shape comparison, the QSAR(II) results and the enhancement of the activity with the introduction of *gem*-dimethyl (metconazole) and isopropyl groups (ipconazole) to the cyclopentane ring indicate that the *gem*-dimethyl and isopropyl groupings play an essential role in binding with P450$_{14DM}$. These structures are possibly assigned as another hydrophobic interaction site.

Poulos *et al*. (*13*) reported the crystal structure of the cytochrome P450 camphor 5-exohydroxylase (P450$_{cam}$) as the camphor bound form. It is difficult to obtain direct information as to the binding mode of metconazole from their results because of a low homology of amino acid sequences between P450$_{14DM}$ and P450$_{cam}$ (*14*). However, the binding site of metconazole in P450$_{14DM}$ is supposed to be surrounded by hydrophobic amino acid residues, similar to the camphor binding site in P450$_{cam}$.

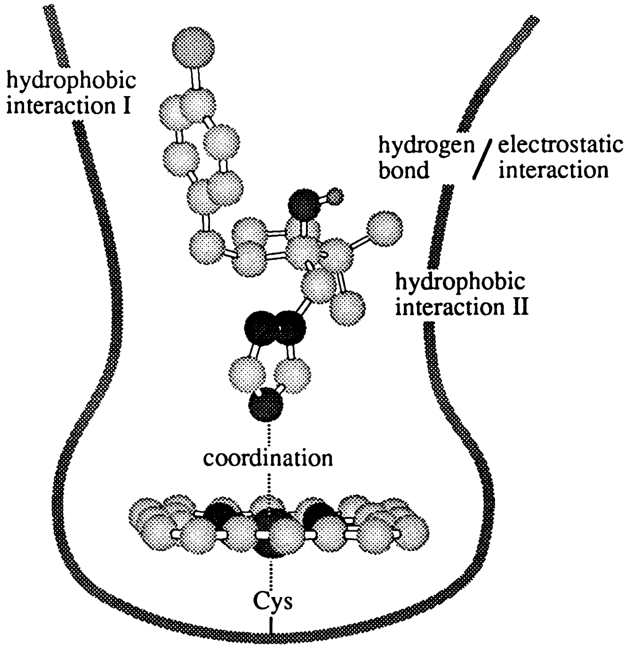

hydrophobic interaction I
hydrogen bond / electrostatic interaction
hydrophobic interaction II
coordination
Cys

Figure 8. Schematic diagram illustrating the binding of (1*S*,5*R*)-metconazole to cytochrome $P450_{14DM}$.

Concluding Remarks

We have been successful in developing the novel potent fungicides of great promise, metconazole (**19**) and ipconazole (**33**). In this paper, we focused on the processes of structural modification from a lead compound to metconazole and ipconazole. During these processes, QSAR analyses of the Hansch-Fujita type and three-dimensional shape comparisons of molecules played critical roles. Furthermore, these procedures have made a great contribution to a deeper understanding of the binding mode of azole type fungicides to the target receptor, $P450_{14DM}$.

Finally, we would like to stress that the discovery of metconazole and ipconazole are from a careful and synergistic cooperation of our synthetic, biological and physical-chemical scientists over a period of several years.

Acknowledgments

We are indebted to Dr. Chizuko Kabuto of Tohoku University for the X-ray measurements. We would like to express our sincere thanks to our many colleagues involved in this special project in Kureha Chemical Industry Co., Ltd., in particular, Drs. Masaaki Takahashi and Yasuo Amagi who have organized this project and led us.

Literature Cited

1. Köller, W. *Pestic. Sci.* **1987**, *18*, 129.
2. Verloop, A.; Hoogenstraaten, W.; Tipker, J. In *Drug Desin Vol. VII* ; Ariëns, E. J. Ed.; Academic Press: New York, **1976**; pp. 165-206.

3. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
4. Leach, A. R. In *Reviews in Computational Chemistry , Volume 2* ; Lipkowitz, K. B.; Boyd, D. B., Ed.; VCH Publishers, Inc.: New York, **1991**; pp. 1-55.
5. Goto, H.; Osawa, E. *J. Am. Chem. Soc.* **1989**, *111*, 8950.
6. Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
7. Funaki, Y.; Ishiguri, Y.; Kato, T.; Tanaka, S. *J. Pesticide Sci.* **1984**, *9*, 229.
8. Funaki, Y.; Ishiguri, Y.; Kato, T.; Tanaka, S. *ibid.* **1983**, *8*, 309.
9. Takano, H. *Jpn. Pestic. Inform.* **1986**, No. 49, 18.
10. Katagi, T. *J. Agric. Food Chem.* **1988**, *36*, 344.
11. Momany, F. A.; Chuman, H. *Methods in Enzym.* **1986**, *124*, 2573.
12. Yoshida, Y.; Aoyama, Y. In *In Vitro and In Vivo Evaluation of Antifungal Agents*; Iwata, K.; Vanden Bossche, H., Eds.; Elsevier Amsterdam : **1986**; pp. 123-134.
13. Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *J. Biol. Chem.* **1985**, *260*, 16122.
14. Ishida, N.; Aoyama, Y.; Hatanaka, R.; Oyama, Y.; Imajo, S. ; Ishiguro, M.; Oshima, T.; Nakazato, H. ; Noguchi, T.; Maitra, U. S.; Mohan, V. P.; Sprinson, D. B.; Yoshida, Y. *Biochem. Biophys. Res. Commun.* **1988**, *155*, 317.

RECEIVED April 26, 1995

Chapter 14

Structure–Activity Relationships of Quinone and Acridone Photosystem II Inhibitors

Wilfried Draber[1], Achim Trebst[2], and Walter Oettmeier[2]

[1]In den Birken 81, 42113 Wuppertal, Germany
[2]Lehrstuhl Biochemie der Pflanzen, Ruhr-Universität, 44780 Bochum, Germany

Photosynthesis inhibition was determined with thylakoids of the green alga *Chlamydomonas reinhardtii* on eleven acridones, twelve 1,4-benzoquinones and fifteen 1,4-naphthoquinones. The wild type and five mutants of the algae were used. The mutants had known changes in the inhibitor-binding region of the D-1 protein. The compounds showed reduced as well as enhanced inhibitory activity up to two orders of magnitude. Their resistance/supersensitivity pattern and structure-activity studies revealed that the inhibitory behaviour resembles that of the nitrophenols.

The photosystem II reaction center is composed of a heterodimer of the D-1 and D-2 proteins. In analogy to the L- and M-subunits of the reaction center from photosynthetic bacteria, they carry the pigments necessary for the primary charge separation and, in addition, the primary and secondary quinone acceptors Q_A and Q_B, respectively. It has been known for a long time, that photosystem II inhibitors, some of which are in use as powerful herbicides, can displace plastoquinone from its Q_B-binding site at the D-1 protein (*1-5*). It has been proven many times that this displacement is competitive (*6*).

Though X-ray crystallography of the photosystem II reaction center with sufficiently high resolution is not available yet, amino acids participating in plastoquinone and inhibitor binding have been identified by analysis of herbicide resistant mutants and photoaffinity labelling (*7, 8*). Thus, the herbicide and plastoquinone binding region within the D-1 protein reaches at least from Met_{211} to Leu_{273} with a total of 63 amino acids. This binding region includes part of transmembrane helix IV, a parallel helix and part of transmembrane helix V (*1-5*). In their mode of binding, two different types of inhibitor families can be distinguished: the Ser_{264}-family, which includes the "classical" herbicides like ureas, triazines, triazinones and carbamates and the His_{215}-family to which belong phenolic herbicides, but also hydroxy-pyridines and -quinolines (*1, 3*). Ser_{264} plays an important role in the binding of "classical" herbicides. It is hydrogen bonded to a carbonyl group of plastoquinone (see Figure 1). If Ser_{264} is mutated to Gly or Ala in herbicide tolerant weeds or algae, the inhibitory potency of

0097–6156/95/0606–0186$12.00/0

"classical" herbicides and inhibitors is reduced. This loss can be as high as three orders of magnitude for certain triazines and triazinones. On the other hand, phenolic inhibitors in these mutants show an increased activity (negative cross resistance, supersensitivity) (*9*). This supersensitivity may amount to more than two orders of magnitude, as is the case for some halogenated benzo- and naphthoquinones (*10*).

Based on the crystallographic data from the reaction center of *Rhodopseudomonas viridis*, models of the Q_B-binding niche have been constructed and various inhibitors were fitted into the binding niche (*11-16*). Like molecular modeling, QSAR may also assist in the optimization of the biological activity of a certain class of inhibitors. During the last ten years about 30 new QSAR of photosystem II inhibitors have been published (*7, 8*). Obviously there is a characteristic difference in parameter sets between "classical" and phenolic inhibitors which are needed to describe their inhibitory activity. Whereas the QSAR of the "classical" inhibitors is governed by the lipophilicity, as expressed by π or log P (*17*), in the QSAR of phenolic inhibitors steric parameters prevail. This was found as early as 1979 (*18*) and confirmed most recently by new data sets (*16, 19*).

Here we wish to report on the inhibitory activity of acridones, 1,4-benzo- and naphthoquinones on electron transport in D-1 protein mutants (Val_{219} > Ile, Ala_{251} > Val, Phe_{255} > Tyr, Ser_{264} > Ala, and Leu_{275} > Phe) of the green alga *Chlamydomonas reinhardtii*. It should be noted that herbicide resistant algae were known as early as 1981(*20*). Figure 1 shows the position of plastoquinone in the binding niche with its hydrogen bonds to Ser264 and His215. The mutations are shown in Figure 1 in black. In these mutants tolerance as well as supersensitivity have been observed. QSAR calculations were performed using either steric or electronic/lipophilicity descriptors. The goal of this investigation is to find an answer to the question into which family acridones, 1,4-benzo- and naphthoquinones have to be grouped.

Materials and Methods

Compounds Studied. Some of the compounds were commercially available. The others have been prepared according to known methods as reported previously (*10*).

The Mutants. Mutants of *Chlamydomonas reinhardtii* (MZ1 = $ser_{264}ala$, MZ2 = $ala_{251}val$, and MZ4 = $leu_{275}phe$) were obtained and grown as decribed previously (*21*). Strains Ar207 ($phe_{255}tyr$) and Dr2 ($val_{219}ile$) were originally isolated by Galloway and Mets (*22*) and were kindly provided by Dr. J. D. Rochaix, University of Geneva, Switzerland.

Thylakoids from *Chlamydomonas reinhardtii* wild type and mutant lines were isolated and assayed for photosystem II activity (DCPIP reduction with water as the electron donor) in the presence of inhibitors according to a previously established protocol (*21*). Most experimental data obtained in this way have been published previously (*10*).

Calculations. Multiple regression analyses were performed on a VAX 7000/610 using the program RS/1 (*23*). The σ and π descriptors were taken from the compilation of Hansch and Leo (*24*), the STERIMOL parameters from Verloop (*25*). Furthermore, we used the F and R values of Hansch, Leo and Taft (*26*).

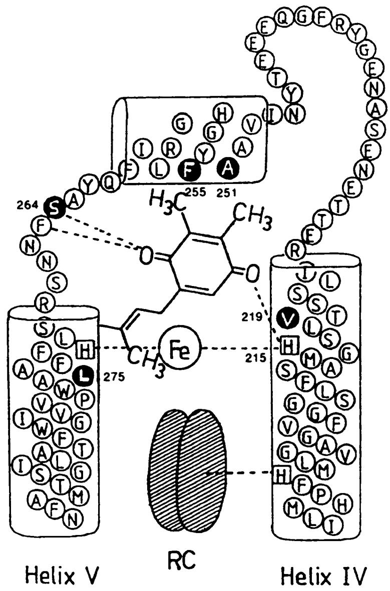


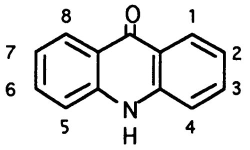
Figure 1. Folding of the amino acid sequence of the D-1 protein in plastoquinone and inhibitor binding. The amino acid substitutions in the five mutants are indicated by black circles. Plastoquinone is hydrogen bonded to His_{215} and Ser_{264}. From ref. 5, with permission.

Results

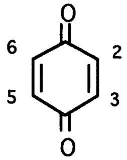
Inhibitory Potency. The substituted quinones and acridones have been tested in cell free thylakoid membrane preparations from the green alga *Chlamydomonas reinhardtii.* Table I lists the structures and the pI_{50} values in wild type and five mutants. pI_{50} is as usual the negative logarithm of the concentration to inhibit photosynthetic electron flow from water to an artificial acceptor (DCPIP) to 50 %. To present a better survey we have listed in Table II the differences of the respective mutant compared to the wild type pI_{50}.

The ΔpI_{50} values present an interesting pattern. In the acridone series, halo-nitro (No.s **8,9,10,11**) derivatives show the highest inhibitory potency. The values in the *Chlamydomonas* systems are very similar to those obtained for some of the compounds in thylakoid preparations from spinach (*27*). An important point is, that most benzoquinones and naphthoquinones become more active inhibitors with many mutants. Especially, they are supersensitive in the $ser_{264}ala$ mutant in contrast to classical inhibitors like diuron, atrazine or metribuzin. This finding suggests that they have to be grouped in the His_{215} family (*2*) like the substituted nitrophenols (*16, 19*). Nine of the 11 acridones, however, display resistance with this mutant. Accordingly, by this criterion the acridones should belong to the Ser_{264} family. Another feature points in the same direction: With two small exceptions, all acridones show enhanced inhibitory activity in the $val_{219}ile$ mutant whereas most benzoquinones become resistant. The greater part of the naphthoquinones show enhancement. Compared with the nitrophenols this observation is less clear. With only three exceptions of the 44 nitrophenolic compounds investigated, all of them displayed decreased activity in this mutant, in

Table I. The Inhibition of Photosynthetic Electron Flow - pI_{50} - by Acridones, Benzoquinones, and Naphthoquinones in Wildtype and Five Mutants of *Chlamydomonas rh.*



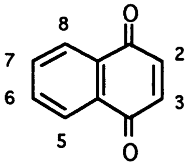
No.	Acridone	w.t.	mutants				
			val 219 ile	ala 251 val	phe 255 tyr	ser 264 ala	leu 275 phe
1	2,4,7-$(NO_2)_3$	5.10	5.92	5.60	6.47	5.60	5.26
2	4-OH	5.33	5.40	5.22	6.15	4.70	5.33
3	2-CMe_3	5.61	6.22	5.70	6.61	4.85	5.31
4	2-CH_2CHMe_2	5.70	6.22	5.92	6.26	5.24	5.70
5	2-F,4-Me,5,7-$(NO_2)_2$	6.00	6.32	6.00	6.42	5.57	5.44
6	2,4,5-$(NO_2)_3$	6.07	6.10	6.03	6.03	5.52	6.08
7	2,4,5,7-$(NO_2)_4$	6.26	6.25	6.42	6.32	6.34	6.54
8	1,4-Cl_2,5,7-$(NO_2)_2$	6.30	6.40	6.46	6.75	6.22	5.95
9	2,4-Cl_2,5,7-$(NO_2)_2$	6.36	6.62	6.32	6.30	6.35	6.15
10	2-Br,4,5,7-$(NO_2)_3$	6.52	6.33	6.46	6.42	6.07	6.60
11	4-Br,2,5,7-$(NO_2)_3$	7.28	7.40	7.06	7.53	7.15	7.53



No.	1,4-Benzoquinone	w.t.	mutants				
			val 219 ile	ala 251 val	phe 255 tyr	ser 264 ala	leu 275 phe
12	2,5,6-Br_3,3-Me	4.44	4.48	4.50	4.60	6.31	5.00
13	2,3-Cl_2,5-CMe_3	4.49	4.85	4.77	4.72	6.05	4.60
14	2,5-Br_2,3-Me,6-$CHMe_2$	4.70	4.85	4.82	4.82	7.18	4.89
15	2,3,5,6-Cl_4	4.74	4.50	4.77	4.77	6.15	5.37
16	2,5-Br_2,3,6-$(OCHMe_2)_2$	4.77	4.70	4.92	4.85	6.60	5.16
17	2,3-Br_2,5-CMe_3	5.53	5.30	4.51	4.72	6.55	4.96
18	2,3,5,6-Br_4	5.82	5.62	5.33	5.80	7.15	6.43
19	2,3,5,6-I_4	6.76	6.59	6.89	7.22	8.08	7.08
20	2,3-I_2,5-CMe_3	6.85	6.37	5.60	6.64	7.19	5.30
21	2,5-$(NO_2)_2$,3-Me,6-$CHMe_2$	4.71	4.85	5.30	4.68	4.89	5.00
22	2,5,6-Br_3,3-$(CH_2)_5Me$	5.24	5.85	5.22	5.35	6.92	5.10
23	2-CMe_3,5,6-Br_2	5.89	5.26	4.48	5.10	6.22	4.77

Continued on next page

Table I. *Continued*

No.	1,4-Naphthoquinone	w.t.	mutants				
			val 219 ile	ala 251 val	phe 255 tyr	ser 264 ala	leu 275 phe
24	2-Br	4.24	4.98	4.37	4.67	5.30	4.39
25	2-Cl,3-Me	5.00	5.58	4.92	5.08	6.15	4.96
26	2-CH$_2$Cl,3-Cl	5.00	5.21	5.38	5.13	6.10	4.71
27	2,3-Cl$_2$	5.33	5.58	5.77	5.46	6.80	5.11
28	2-Br,3-Me	5.38	5.59	5.19	5.51	6.05	5.60
29	2,3-Br$_2$	5.76	6.16	6.40	5.87	7.20	5.31
30	2-Br,3-(CH$_2$)$_6$Me	5.80	5.89	5.85	6.01	6.40	5.77
31	2-Br,3-CHMe$_2$	5.89	5.68	5.77	5.71	6.40	5.68
32	2-Br,3-CH$_2$Ph	6.00	5.90	6.10	6.10	6.19	5.91
33	2,3-I$_2$	6.10	6.43	6.36	6.16	7.19	6.30
34	2-(CH$_2$)$_3$Me	4.74	4.74	4.68	4.52	4.44	4.68
35	2-CH$_2$Ph	4.82	5.05	4.70	4.92	4.74	4.90
36	2-(CH$_2$)$_6$Me	4.60	4.60	4.51	4.85	4.48	4.17
37	2,3-(CN)$_2$	5.03	5.21	4.85	5.30	5.05	5.27
38	2-CHMe$_2$,3-Br	5.39	5.34	5.38	5.36	6.13	5.22

part to a great degree (ΔpI_{50} up to +1.7, 2-bromo-4-nitro-6-benzylphenol) compared to a meagre +0.63 for 2-*t*-butyl-5,6-dibromo-benzoquinone (No. 23). This reflects that the acridones, benzo- and naphthoquinones must have positions different from the nitrophenols in the Q_B binding niche. Moreover, one may speculate that the benzoquinones are bound in the binding niche after reduction to hydroquinones. The compounds of the present study show a mixed behaviour towards the ala$_{251}$val and leu$_{275}$phe mutants whereas in the phe$_{255}$tyr mutant enhanced sensitivity prevails. This differs from the nitrophenols as they display decreased sensitivity in most cases with this mutant.

The Data Base. The compounds were not selected for statistical treatment and many of them were reported earlier (*10*). Therefore, the composition of the sample is rather biased. This holds true especially for the acridones where 8 of the 11 compounds are substituted with two or three nitro groups. In the benzo- and naphthoquinones, electron attracting and donating substituents are more evenly distributed, but still the small number of compounds is critical. This has to be kept in mind when QSAR results are interpreted.

A too small spread of the biological data can lead to statistically ambiguous results. As a prerequisite for calculating the regression equations we have checked the range of pI_{50} values in wild type and the five mutant chloroplasts for acridones, benzoquinones, and naphthoquinones. They are given in Table III.

Table II. Values of ΔpI_{50} [= $pI_{50w.t.}$ - $pI_{50mutant}$]

1. Acridones

No.	w.t.	val 219 ile	ΔpI_{50}	ala 251 val	ΔpI_{50}	phe 255 tyr	ΔpI_{50}	ser 264 ala	ΔpI_{50}	leu 275 phe	ΔpI_{50}
1	5.10	5.92	-0.82	5.60	-0.50	6.47	-1.37	5.60	-0.50	5.26	-0.16
2	5.33	5.40	-0.07	5.22	+0.11	6.15	-0.82	4.70	+0.63	5.33	0.00
3	5.61	6.22	-0.61	5.70	-0.09	6.61	-1.00	4.85	+0.76	5.31	+0.30
4	5.70	6.22	-0.52	5.92	-0.22	6.26	-0.56	5.24	+0.46	5.70	0.00
5	6.00	6.32	-0.32	6.00	0.00	6.42	-0.42	5.57	+0.43	5.44	+0.56
6	6.07	6.10	-0.03	6.03	+0.04	6.03	+0.04	5.52	+0.55	6.08	-0.01
7	6.26	6.25	+0.01	6.42	-0.16	6.32	-0.06	6.34	-0.08	6.54	-0.28
8	6.30	6.40	-0.10	6.46	-0.16	6.75	-0.45	6.22	+0.08	5.95	+0.35
9	6.36	6.62	-0.26	6.32	+0.04	6.30	+0.06	6.35	+0.01	6.15	+0.21
10	6.52	6.33	+0.19	6.46	+0.06	6.42	+0.10	6.07	+0.45	6.60	-0.08
11	7.28	7.40	-0.12	7.06	+0.22	7.53	-0.25	7.15	+0.13	7.53	-0.25
2. Benzoquinones											
12	4.44	4.48	-0.04	4.50	-0.06	4.60	-0.16	6.31	-1.87	5.00	-0.56
13	4.49	4.85	-0.36	4.77	-0.28	4.72	-0.23	6.05	-1.56	4.60	-0.11
14	4.70	4.85	-0.15	4.82	-0.12	4.82	-0.12	7.18	-2.48	4.89	-0.19
15	4.74	4.50	+0.24	4.77	-0.03	4.77	-0.03	6.15	-1.41	5.37	-0.63
16	4.77	4.70	+0.07	4.92	-0.15	4.85	-0.08	6.60	-1.83	5.16	-0.39
17	5.53	5.30	+0.23	4.51	+1.02	4.72	+0.81	6.55	-1.02	4.96	+0.57
18	5.92	5.62	+0.30	5.33	+0.59	5.80	+0.12	7.15	-1.23	6.43	-0.51
19	6.76	6.59	+0.17	6.89	-0.13	7.22	-0.46	8.08	-1.32	7.08	-0.32
20	6.85	6.37	+0.48	5.60	+1.25	6.64	+0.21	7.19	-0.34	5.30	+1.55
21	4.71	4.85	-0.14	5.30	-0.59	4.68	+0.03	4.89	-0.18	5.00	-0.29
22	5.24	5.85	-0.61	5.22	+0.02	5.35	-0.11	6.92	-1.68	5.10	+0.14
23	5.89	5.26	+0.63	4.48	+1.41	5.10	+0.79	6.22	-0.33	4.77	+1.12

Continued on next page

Table II. *Continued*

3. Naphthoquinones

No.	w.t.	val 219 ile	ΔpI_{50}	ala 251 val	ΔpI_{50}	phe 255 tyr	ΔpI_{50}	ser 264 ala	ΔpI_{50}	leu 275 phe	ΔpI_{50}
24	4.24	4.98	-0.74	4.37	-0.13	4.67	-0.43	5.30	-1.06	4.39	-0.15
25	5.00	5.58	-0.58	4.92	+0.08	5.08	-0.08	6.16	-1.16	4.96	+0.04
26	5.00	5.21	-0.21	5.38	-0.38	5.13	-0.13	6.10	-1.10	4.71	+0.29
27	5.33	5.58	-0.25	5.77	-0.44	5.46	-0.13	6.80	-1.47	5.11	+0.22
28	5.38	5.59	-0.21	5.19	+0.19	5.51	-0.13	6.05	-0.67	5.60	-0.22
29	5.76	6.16	-0.40	6.40	-0.64	5.87	-0.11	7.20	-1.44	5.31	+0.45
30	5.80	5.89	-0.09	5.85	-0.05	6.01	-0.21	6.40	-0.60	5.77	+0.03
31	5.89	5.68	+0.21	5.77	+0.12	5.71	+0.18	6.40	-0.51	5.68	+0.21
32	6.00	5.90	+0.10	6.10	-0.10	6.10	-0.10	6.19	-0.19	5.91	+0.09
33	6.10	6.43	-0.33	6.36	-0.26	6.16	-0.06	7.19	-1.09	6.30	-0.20
34	4.74	4.74	0.00	4.68	+0.06	4.52	+0.22	4.44	+0.30	4.68	+0.06
35	4.82	5.05	-0.23	4.70	+0.12	4.92	-0.10	4.74	+0.08	4.90	-0.08
36	4.60	4.60	0.00	4.51	+0.09	4.85	-0.25	4.48	+0.12	4.17	+0.43
37	5.03	5.21	-0.18	4.85	+0.18	5.30	-0.27	5.05	-0.02	5.27	-0.24
38	5.39	5.34	+0.05	5.38	+0.01	5.36	+0.03	6.13	-0.74	5.22	+0.17

Based on Table III, we report calculations employing pI_{50} values of wild type and the mutants $ser_{264}ala$ and $leu_{275}phe$ of all compound types. Further, regression calculations were performed on acridones and benzoquinones with pI_{50} values of the $val_{219}ile$ and the $ala_{251}val$ mutant, and on naphthoquinones with the $ala_{251}val$ mutant. With the $phe_{255}tyr$ mutant, for insufficient significance no regressions are reported .

Quantitative Structure Activity Relationships. The best equations obtained from the data base are given in Tables IV (acridones), V (benzoquinones), and VI (naphthoquinones). In these equations F, r^2 and s have the usual meaning (F-test of the regression; squared correlation coefficient; standard deviation). The numbers in parentheses below the coefficients are the corresponding t-test values. In order to judge the relationship of the standard deviation to range of the respective sample, the ratio $s/\Delta pI_{50}$ was calculated. If this figure exceeds 0.15, the significance of the equation is to be regarded with caution.

Discussion of the Regression Equations. One of the objectives of the investigation was to determine whether the inhibitors reported here - acridones, benzoquinones, naphthoquinones - are members of either the Ser264 (or classical) or the His215 (or phenol like) family which has implications for their

Table III. Maximum Range of pI_{50} Values

	Mutant	Most Active Compound	Least Active Compound	ΔpI_{50}
1. Acridones, n = 11				
	wild type	11	1	2.18
	$val_{219}ile$	11	2	2.00
	$ala_{251}val$	11	2	1.84
	$phe_{255}tyr$	11	6	1.50
	$ser_{264}ala$	11	2	2.45
	$leu_{275}phe$	11	3	2.27
2. Benzoquinones, n = 12				
	wild type	20	12	2.41
	$val_{219}ile$	19	12	2.11
	$ala_{251}val$	19	23	2.41
	$phe_{255}tyr$	19	23	2.62
	$ser_{264}ala$	19	21	3.19
	$leu_{275}phe$	19	13	2.48
3. 1,4-Naphthoquinones, n = 15				
	wild type	33	24	1.86
	$val_{219}ile$	33	36	1.83
	$ala_{251}val$	33	24	1.99
	$phe_{255}tyr$	33	34	1.64
	$ser_{264}ala$	29	34	2.76
	$leu_{275}phe$	33	36	2.13

Table IV. Regression Equations of 11 Acridones

Equ.	**Mutant**	pI_{50} =					Intercept	F	r^2	s	s/ΔpI_{50}
1	**wild type**	L_2 +0.29 (+2.05	L_4 +0.62 +2.77	L_7 -0.72 -2.71	$B5_7$ +1.00 +4.63)		+3.47	9.8	0.88	0.29	0.13
2	**val$_{219}$ile**	$B1_4$ - 0.27 (- 1.41	$B5_4$ - 1.31 - 3.61	$\Sigma\sigma$(CO) + 0.83 + 4.54)			+ 7.87	7.7	0.77	0.28	0.14
3	**ala$_{251}$val**	$B5_2$ + 0.19 (+ 1.89	L_5 + 0.69 + 4.65)				+ 3.66	10.8	0.73	0.29	0.16
4	**ser$_{264}$ala**	$B5_2$ + 0.36 (+ 7.14	L_4 + 1.04 - 5.42	$B5_4$ - 0.70 + 4.83	$B5_5$ + 0.41 + 5.49	L_7 + 0.49 + 7.01)	+ 0.82	57.4	0.98	0.13	0.05
5	**leu$_{275}$phe**	$B5_2$ + 0.68 (+ 3.31	$B1_4$ + 1.10 + 2.72	L_5 + 0.99 + 4.87)			- 0.30	8.5	0.78	0.39	0.17

Table V. Regression Equations of 12 Benzoquinones

Equ.	**Mutant**	pI_{50} =				Intercept	F	r^2	s	s/ΔpI_{50}
6	**wild type**	$B1_2$ + 2.23 (+ 7.76	$B1_3$ + 3.08 + 10.96	$B5_3$ + 0.20 + 3.19	$B5_6$ + 0.33 + 3.21)	- 5.60	35.7	0.95	0.23	0.10
7	**val$_{219}$ile**	L_2 + 8.24 (+ 4.25	$B1_2$ - 13.87 - 3.81	L_3 + 0.24 + 5.86	$B1_3$ + 1.57 + 8.96)	- 2.95	36.8	0.95	0.19	0.09
8	**ala$_{251}$val**	L_2 + 1.23 (+ 3.79	$B1_3$ + 1.51 + 2.95	$B1_5$ - 1.10 - 2.05)		+ 0.17	5.0	0.65	0.46	0.19
9	**ser$_{264}$ala**	$B1_2$ + 4.21 (+ 8.13	$B5_2$ - 2.14 - 4.63	$B1_3$ + 1.63 + 4.61	$B5_6$ + 0.49 + 3.71)	- 1.09	18.7	0.91	0.29	0.14
10	**leu$_{275}$phe**	L_2 + 1.19 (+ 8.86	$B1_3$ + 3.06 + 12.43	$B5_3$ + 0.19 + 3.64	$B5_6$ + 0.30 + 3.30)	- 5.68	46.1	0.96	0.20	0.08

Table VI. Regression Equations of 15 Naphthoquinones

Equ.	**Mutant**	pI_{50} =	Intercept	F	r^2	s	$s/\Delta pI_{50}$
11	**wild type**	$B1_3$ $B5_3$ + 0.91 + 0.17 (+ 5.29 + 4.32)	+ 3.47	30.9	0.84	0.24	0.13
12	**$ala_{251}val$**	$B1_3$ $B5_3$ + 1.22 + 0.15 (+ 5.43 + 3.01)	+ 3.09	24.9	0.81	0.32	0.16
13	**$ser_{264}ala$**	$B1_2$ $B1_3$ + 1.73 + 1.49 (+ 3.13 + 4.98)	+ 0.48	38.1	0.86	0.36	0.13
14	**$leu_{275}phe$**	$B1_3$ $B5_3$ + 0.82 + 0.18 (+ 3.28 + 3.03)	+ 3.52	13.3	0.69	0.35	0.19

positioning in the binding niche of D-1. The regression equations can assist in reaching this goal. This should be in line with Hansch's endeavours (*28*) in drawing general conclusions from an enormous data base of regression equations on diverse types of compounds.

Only the best equations, according to the statistical criteria, are given. Of the tested descriptors σ, π, F, R and STERIMOL, the latter ones resulted in the statistically best regression equations with one exception. This is Equation 2 (for the val219ile mutant) in which the descriptor $\Sigma\sigma(CO)$ stands for an addition of all σ_m and σ_p values that affect the carbonyl group in the acridones ($\Sigma\sigma(CO)$ = $\Sigma(\sigma_{p1} + \sigma_{m2} + \sigma_{p3} + \sigma_{m4} + \sigma_{m5} + \sigma_{p6} + \sigma_{m7} + \sigma_{p8})$. F and r^2 of Equation 2, however, show only moderate significance. A similar procedure for the NH-group gave negative results for all mutants. Thus it can be stated, that generally the STERIMOL parameters lead to the most significant equations, though there is no lack of electron attracting substituents. This resembles the investigations we carried out with nitrophenols (*16, 19*). They belong in the His_{215} family for biochemical reasons. QSAR and molecular modelling studies point in the same direction. Equation 2 leaves some uncertainty as to the grouping of acridones into the His_{215} family. Similar doubts arise from the ΔpI_{50} values with regard to the $ser_{264}ala$ mutant of Table III. However, the extremely biased acridone sample demands caution.

Another feature of the equations is that the signs of the STERIMOL parameters with a few exceptions (Equations 2, 4, 7, 9, in each only one of three or four descriptors) are positive. That means that the contribution of the various amino acid side chains and their polar groups with the inhibitors is positive. Another point deserves attention. All the mutations mean an increase in lipophilicity of the changed amino acids, the most significant one being the mutation of serine to alanine. Therefore, it appears justified to call the interaction between inhibitor and the amino acids of the binding niche lipophilic in nature, however, with strong steric restrictions.

Conclusion

The majority of benzo- and naphthoquinones exhibit enhanced inhibitory activity in the $ser_{264}ala$ mutant. Also the calculations have confirmed that these compounds have to be grouped into the His_{215} family. As to the acridones, the same conclusion can not be drawn, neither from their behaviour in comparison to the wild type pI_{50} values, nor as a result from the regression equations. However, there is other evidence for their grouping as members of the His_{215} family as well (*10*).

The importance of STERIMOL parameters in the regression equations suggests that the orientation of the compounds in the Q_B binding niche is mainly by hydrophobic interaction that is very sensitive to steric restrictions of certain amino acid side chains, in spite of presence of many polar substituents in the inhibitor molecules. Molecular modelling studies with the nitrophenols (*16*) have already pointed in this direction. In constrast to that, the interaction of the Ser_{264}-family inhibitors (*2*) with the binding niche is through hydrophobic *and* electrostatic forces.

Acknowledgements

W. Draber is indebted to Bayer AG Leverkusen for the opportunity to use their soft- and hardware.

References

1. Trebst, A.; Draber, W. *Photosyn. Res.* **1986**, *10*, 381.
2. Trebst, A. *Z. Naturforsch.* **1986**, *41c*, 240.
3. Trebst, A. *Z. Naturforsch.* **1987**, *42c*, 742.
4. Trebst, A.; Kluth, J.; Tietjen, K.; Draber, W. *Proc. of the 7th Int. Congress of Pesticide Chemistry*, H. Frehse Ed., VCH: Weinheim, 1990, pp.111-120.
5. Draber, W.; Tietjen, K.; Kluth, J. F.; Trebst, A. *Angew. Chem. Int .Ed.* **1991**, *30*, 1621.
6. Oettmeier, W.; Soll, H.J. *Biochim. Biophys. Acta* **1983**, *724*, 287.
7. Oettmeier, W. *Topics in Photosynthesis*, Barber J. Ed., Elsevier: Amsterdam, London, New York, Tokyo, l992, Vol. 11, 349-408.
8. Bowyer, J.R.; Camillieri, P.; Vermaas, W.F.J. *Topics in Photosynthesis*; Baker, N. R., Percival M. P. Eds.; Elsevier: Amsterdam, London, New York, Tokyo, 1991, Vol. 10, pp. 27-85.
9. Oettmeier, W.; Masson, K.; Fedtke, C.; Konze, J.; Schmidt, R.R. *Pestic. Biochem. Physiol.* **1982**, *18*, 357.
10. Oettmeier, W.; Masson, K.; Kloos, R.; Reil, E. *Z. Naturforsch.* **1993**, *48c*, 146.
11. Bowyer, J.; Hilton, M.; Whitelegge, J.; Jewess, P.; Camilleri, P.; Crofts, A.; Robinson, H.; *Z. Naturforsch.* **1990**, *45c*, 379.
12. Tietjen, K.G.; Kluth, J.F.; Andree, R.; Haug, M.; Lindig, M.; Mueller, K.H.; Wroblowsky, H.J.; Trebst, A. *Pest. Sci.* **1990**,*31*, 65.
13. Mackay, S.P.; O'Malley, P J. *Z. Naturforsch.* **1993**,*48c*,474.
14. Mackay, S.P.; O'Malley, P.J. *Z. Naturforsch.* **1993**,*48c*,773.
15. Mackay, S.P.; O'Malley, P.J. *Z. Naturforsch.* **1993**,*48c*,782,.
16. Draber, W.; Hilp, U.; Schindler, M.; Trebst, A., *Natural and Engineered Pest Management Agents*; Hedin, P.A.; Menn, J. J.; Hollingworth, R., Eds.; ACS Symp. Series 551: Washington D.C., 1994; 449-472.
17. Kakkis, E.; Palmire, V.C.; Strong, C.D.; Bertsch,W.; Hansch, C.; Schirmer, U. *J. Agric. Food Chem.* **1984**, *32*, l33.

18. Trebst, A.; Draber,W. *Adv. in Pesticide Science* ; Geissbühler H., Ed.; Pergamon Press: Oxford, New York, 1979, Vol. 2; pp. 223-234.
19. Trebst, A.; Hilp. U.; Draber, W.; *Phytochem.* **1993**, *33*, 969.
20. Böger, P., Sandmann, G., Miller, R. *Photosyn. Res.* **1981**, *2*, 61.
21. Wildner, G.F.; Heisterkamp, U.; Trebst, A. *Z. Naturforsch.* **1990**, *45c*, 1142.
22. Galloway, R. E.; Mets, L. *J. Plant Physiol.* **1984**,*74*, 469.
23. BBN Software Products Corporation, Cambridge, Massachusetts 02238.
24. Hansch, C.; Leo, A. *Substitution Constants for Correlation Analysis in Chemistry and Biology*, Wiley: New York, 1979.
25. Verloop, A. *Pesticide Chemistry, Human Welfare and the Environment*, Miyamaoto, J. and Kearney, P.C., Ed.s, Pergamon Press: Oxford, 1983, Vol. 1, *The Sterimol Approach: Further Development of the Method and New Applications*, pp. 339-344.
26. Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165-195.
27. Oettmeier, W.: Masson, K.; Kloos, R.; Godde, D. *Research in Photosynthesis*; Murata, N., Ed., Kluwer Academic Publishers: Dordrecht, 1992, Vol III; *The acridones, new inhibitors of electron transport through photosystem II*, pp. 563-566.
28. Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147-153.

RECEIVED July 28, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

Chapter 15

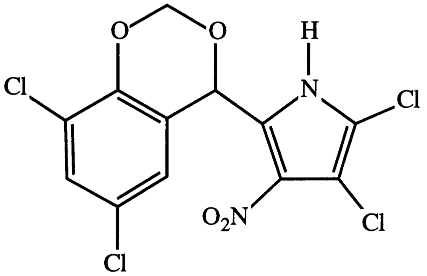
The QSAR of Insecticidal Uncouplers

David M. Gange, Stephen Donovan, Ronald J. Lopata, and Kevin Henegar

**Agricultural Research Division, American Cyanamid Company,
P.O. Box 400, Princeton, NJ 08543–0400**

The discovery that the antibiotic dioxapyrrolomycin possesed potent insecticidal activity led to an extensive study of the Quantitative Structure Activity Relationship (QSAR) of the insecticidal pyrroles. Both mitochondrial assays and *in vivo* studies indicate that uncoupling of oxidative phosphorylation is the primary mode of action of dioxapyrrolomycin and related pyrroles. Uncouplers inhibit the synthesis of ATP, leading to energy starvation and eventual insect death. Our studies indicate that the activity of the pyrrole insecticides is primarily a function of LogP and pK_a. The pyrrole QSAR model allowed us to design and synthesize a new class of insecticidally active uncouplers based on the indole nucleus.

In 1985 scientists at Lederle Laboratories, the pharmaceutical division of American Cyanamid, discovered a new antibiotic, dioxapyrrolomycin. The compound was produced by an unusual species of streptomyces and it inhibited the growth of Gram positive bacteria, Gram negative bacteria, and fungi (*1*). During the course of screening experiments dioxapyrrolomycin was discovered to possess potent *in vivo* insecticidal activity. The Southern Armyworm (*Spodoptera eridania*), Tobacco Budworm (*Heliothis virescens*), Spotted Mite (*Tetranychus urticae*), and Western Potato Leafhopper (*Empoasca abrupta*) were all susceptible to the compound in a leaf



Dioxapyrrolomycin

0097–6156/95/0606–0199$12.00/0

dip assay. The mode of action of dioxapyrrolomycin and related compounds was believed to be uncoupling of oxidative phosphorylation. Our suspicions were confirmed when Bruce Black and Christine Kukel of our laboratories were able to show that dioxapyrrolomycin and related pyrroles acted as uncouplers of rat liver mitochondria *in vitro*.

Background of Uncoupling

The Discovery of Uncoupling. In 1885 Cazeneuve and Lepine administered common food dyes to dogs (2). They reported that one of these dyes, 2,4-dinitro-1-naphthol, exhibited "une assez grande toxicite". The symptoms included an increased breathing rate and a sharp rise in temperature, reaching 44 °C within minutes. We now know that the effects that Cazeneuve and Lepine observed are due to uncoupling.

The first use of an uncoupler as a pesticide dates to 1892 when the potassium salt of 6-methyl-2,4-dinitrophenol was used to control the Nun moth in Europe (*3*). About this time the sodium salt was also used as a herbicide. During the first world war workers who were exposed to dinitrophenol in munitions plants suffered toxic effects including weight loss (*4*). During the 1920's dinitrophenol was given to obese patients to induce weight loss. Unfortunately the treatment led to several deaths as reported in the Journal of the American Medical Association in 1929 (*5*). The use of uncouplers for control of obesity was subsequently discontinued. In 1948 William F. Loomis and Fritz A. Lippman of the Harvard Medical School showed that dinitrophenol inhibits ATP synthesis yet stimulates respiration, the classic definition of uncoupling (*6*). In normally functioning cells respiration and ATP synthesis are tightly coupled. Uncouplers are compounds that uncouple phosphorylation from respiration.

The Mode-of-Action of Uncouplers. In 1961 Mitchell proposed the chemiosmotic hypothesis to explain the mechanism of uncoupling (*7,8*). Mitchell proposed that ATP was synthesized by mitochondria as follows (Figure 1): Oxidation drives protons across the membrane and out of the organelle creating a proton gradient and an electric field. The gradient then drives the protons back into the organelle through the membrane bound catalytic system where they are used to synthesize ATP. In the absence of a proton gradient no ATP is synthesized.

Generally, uncouplers are highly lipophilic weak acids which disrupt the functioning of the mitochondrial membrane. Because uncouplers are highly lipophilic they have a high affinity for lipophilic membranes, such as the mitochondrial membrane. Because uncouplers are weakly acidic they can act as proton shuttles, transporting protons across the membrane. Uncouplers eliminate the proton gradient across the mitochondrial membrane by providing an alternate entry path for protons. They allow protons to enter the organelle without driving the catalytic system. Thus in uncoupled mitochondria oxidation proceeds apace, driving protons out of the mitochondria and generating heat, but no ATP is synthesized, no energy is stored, because the uncoupler short-circuits the catalytic system.

In theory lipophilic weak bases could also act as uncouplers, but in practice weak base uncouplers are rarely observed. Weak bases are probably more susceptible to metabolic degradation, since they are often easily oxidized.

QSAR studies of uncoupling

A number of workers have studied the quantitative structure-activity relationships of uncouplers. Hansch, studying a series of phenol derivatives, found that the inhibition of phosphorylation in yeast was dependent upon two factors, the pK_a and π, the lipophilicity parameter. Activity is improved by increasing acidity and lipophilicity (*9*).

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

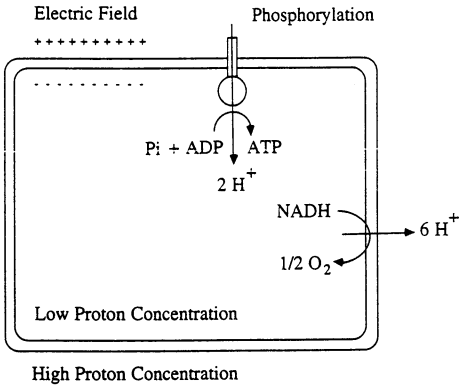
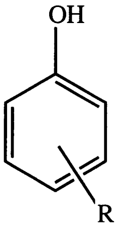

Electric Field
Phosphorylation
+ + + + + + + + + +
- - - - - - - - - -
Pi + ADP
ATP
2 H+
NADH
6 H+
1/2 O2
Low Proton Concentration
High Proton Concentration

Figure 1. Mitochondrial ATP Synthesis.



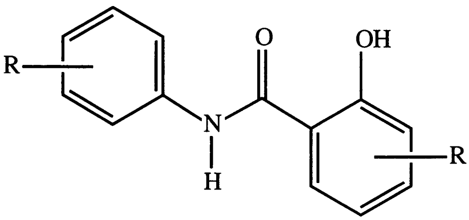
$$\log(1/C) = -0.491 * pK_a + 0.620 * \pi + 6.792 \quad (1)$$
$$n=14 \quad r=0.936 \quad s=0.406$$

The action of phenols on rat liver mitochondria gave the following equation:

$$\log(1/C) = -0.510 * pK_a + 0.400 * \pi + 6.607 \quad (2)$$
$$n=6 \quad r=0.99 \quad s=0.039$$

Again Hansch finds that pK_a and π are the most important factors. It is also interesting to note how closely the two equations resemble one another.

Tollenaere studied the effect of three classes of uncouplers, salicylanilides, 2-trifluoro- methylbenzimidazoles, and phenols, on rat liver mitochondria (*10*). The phenol results were similar to Hansch's results. For the salicylanilides Tollenaere found the following equation :

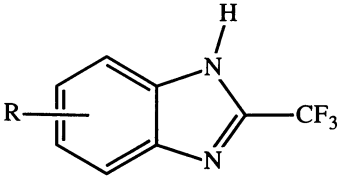


$$pI_{50} = 0.452 * \Sigma\pi + 0.951 * \Sigma\sigma + 5.027 \quad (3)$$
$$n=12 \quad s=0.349 \quad r=0.955 \quad F=47.14$$

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

The important factors are π and σ, and the positive signs of the terms indicate that potency is increased with increasing lipophilicity and acidity. A π^2 term was also calculated for these data and was found to be significant at the 90% level. Thus there seems to be an optimum π value but the data are not extensive enough to allow the optimum to be calculated with accuracy.

In the case of the 2-trifluoromethylbenzimidazoles the following equations were developed:

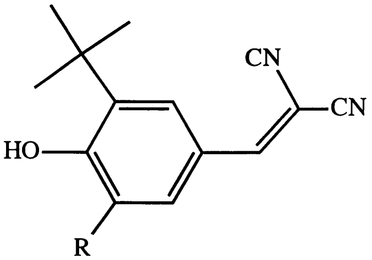


$$pC = 0.580^*\Sigma\pi + 3.104^*\Sigma\sigma - 0.680^*\Sigma\sigma^2 + 3.839 \qquad (4)$$
$$n=22 \quad s=0.531 \quad r=0.951 \quad F=82.98$$

$$pC = 0.421^*\Sigma\pi + 1.641^*pK_a - 0.162^*pK_a{}^2 + 3.560 \qquad (5)$$
$$n=22 \quad s=0.338 \quad r=0.981 \quad F=158.8$$

pC is the log concentration of compound causing maximal stimulation of respiration in the mitochondria. In the case of the 2-trifluoromethylbenzimidazoles the pK_a range tested is wide enough to find an optimum. Differentiation of Equation 5 with respect to pK_a, and solving the resulting equation yields a pK_a optimum of 5.06. Either the sum of sigma values or the pK_a itself could be used to describe the data, although a better explanation is derived using the pK_a values. The important factors are again the lipophilicity, expressed as π, and the pK_a, used directly or expressed in terms of the sum of the sigma values of the substituents.

Studying the effect of phenols on chloroplasts, Miyoshi and Fujita found that the uncoupling activity was related to three factors, the lipophilicity, the acidity, and the steric effect of substituents adjacent to the phenol hydroxyl group (*11*). The steric factor was correlated using the E_s parameter of Taft. Thus:



$$\log(1/C) = 1.070^*\text{LogP} + 1.191^*pK_a - 0.172^*\Sigma E_s{}^{\text{ortho}} + 18.287 \qquad (6)$$
$$n=13 \quad s=0.285 \quad r=0.976$$

The negative sign in front of the E_s term indicates that uncoupling activity is increased by increasing the bulk of the ortho substituents. As before activity is also improved by increasing lipophilicity and acidity.

The preceding studies, and others indicate that uncoupling activity is related to three important factors; the lipophilicity of the compounds, their acidity, and the amount

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

of steric bulk near the acidic site in the compound. In addition most active uncouplers have an extended pi aromatic system which is capable of delocalizing the negative charge of the anion. This delocalization may act to minimize the repulsive interaction between the ionic form of the uncoupler and the mitochondrial membrane. Intra and intermolecular hydrogen bonding are also believed to play a role in shielding the membrane from the full charge of the ionic form of the uncoupler. Across several classes of compounds, including dinitrophenols, 4-dicyanovinylphenols, 2-triflouromethylbenzimidazoles, benzotriazoles, carbonyl cyanide phenylhydrazones, salicylanilides, diarylamines, and triazoles the dependence upon these properties holds (Figure 2) (*12-21*).

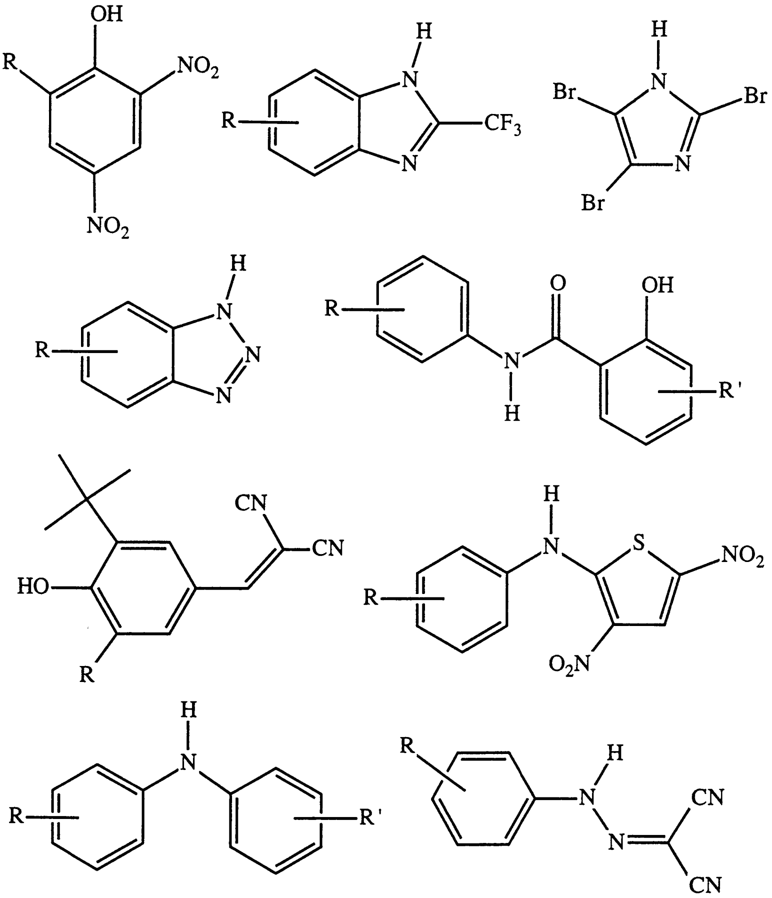


Figure 2. The structures of known classes of uncouplers.

Thus it appears that uncoupling is purely a function of the physical properties of the molecules (*22*).

The Analysis of the Pyrrole Insecticides

With the literature background on uncoupling in mind we set out to study the QSAR of the pyrrole insecticides. Preliminary work indicated that the pyrroles behaved similarly to other uncouplers. The activity was primarily dependent upon the LogP and pK_a of

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

the compounds. Twenty four compounds were selected for study. These compounds were selected so as to cover as broad an area of the important parameter space as possible. While an adequate pK_a correlation could be obtained through the use of the Hammett σ parameter, we decided to measure the pK_a's of the compounds used in the present study. Because of solubility problems the pK_a's were measured in acetone/water solution. The results of the acetone/water measurements were then extrapolated to give pK_a values for pure water. While the extrapolated pK_a's are self consistent, we find that the values are approximately one pK_a unit too high when compared to measurements made in pure water. The LogP values used in the study were generated with the program CLOGP (*23,24*). The biological activity was measured *in vivo* on the Southern Armyworm (*Spodoptera eridania*) in a leaf dip assay. Activities used in developing the models are expressed as Log(1/C) where C is the molar LC_{50} concentration. The numbering scheme used is shown below. The data for the compounds are shown in Table I. The LC_{50} values in Table 1 are expressed in terms of parts per million.

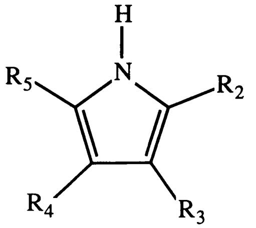


Table I. Pyrrole Data

R2	R3	R4	R5	pK_a	LogP	LC_{50}
Br	Br	NO2	Ph	7.0	4.41	45.00
Br	Br	CN	4Cl Ph	7.4	5.14	3.60
Cl	Cl	CN	4Cl Ph	7.0	5.12	5.10
Cl	Cl	CN	PhSO2CH3	6.5	2.65	315.30
Cl	Br	CN	4Cl Ph	7.1	4.99	4.04
Cl	Cl	CN	2,5Cl Ph	6.3	5.84	3.90
CN	H	CN	3,4Cl Ph	6.8	3.95	9.50
Br	Br	CN	4OCF3 Ph	7.4	5.50	3.10
CN	H	CN	4Cl Ph	8.0	3.23	5.30
CN	Br	Br	3,4Cl Ph	8.5	5.86	3.00
Br	Br	NO2	Br	4.1	2.93	3000.00
Br	NO2	Br	CN	3.7	1.94	737.28
Br	CN	Br	CN	4.4	2.21	132.90
Br	CN	Br	CN	4.4	2.21	196.78
CF3	H	NO2	3,4Cl Ph	7.3	5.60	1.17
CF3	Br	NO2	3,4Cl Ph	5.9	6.23	10.00
CN	4Cl Ph	CN	H	8.1	3.02	51.26
CF3	H	NO2	4Cl Ph	8.1	4.88	5.83
Br	NO2	Br	COPhCl	4.9	4.06	900.00
CF3	Br	2,3Cl Ph	CN	6.2	6.03	10.27
CF3	Br	CN	OC2F5	7.2	6.64	3.00
CF3	4Cl Ph	CN	4Cl Ph	8.0	7.22	6.38
CF3	4Cl Ph	NO2	4Cl Ph	7.5	7.49	3.90
CF3	4Cl Ph	NO2	3,4Cl Ph	7.1	8.21	1.54
CF3	CF3	H	3,4Cl Ph	9.3	6.65	300.00

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

The Development of a pK_a Model for Pyrroles. The development of a model describing the pK_a of pyrroles was our first goal since the activity is most sensitive to changes in pK_a. We wished to help our chemists avoid synthesizing obviously poor compounds. Also if the model was sufficiently accurate it would allow us to calculate, rather than measure, pyrrole pK_a's for QSAR purposes.

Early work by Barlin and Perrin on the prediction of organic acid strengths gave the following formula for substituted azoles, pyrroles, imidazoles, pyrazoles, and purines (*25*).

$$pK_a = 17.0 - 4.28 * \Sigma \sigma \qquad (7)$$

In Equation 7, σ stands for the Hammet sigma parameter, used to estimate substituent electron withdrawing power. The $\Sigma \sigma$ represents the sum of the sigmas of the substituents attached to, and heteroatoms contained within, each ring. Barlin and Perrin claimed a standard deviation of ± 0.4 pH unit. In our hands the model failed to give accurate predictions for pyrrole pK_a's.

Because the prediction of pyrrole pK_a using Equation 7 was problematic a study of pyrrole pK_a's was undertaken. As previously mentioned twenty four compounds were selected for analysis and their pK_a's were measured in an acetone/water solution. Equations 8 and 9 were developed with the data listed in Table I. Equation 8 is based only upon the measured data for the pyrroles. Equation 9 includes the pK_a value for pyrrole itself in the data.

$$pK_a = 15.25 - 7.14 *\Sigma\ \sigma_m(2,5) - 5.07*\Sigma\sigma_p(3,4) \qquad (8)$$
$$n=24 \quad s=0.431 \quad r^2 =0.913 \quad F=110.6$$

$$pK_a = 16.37 - 7.99*\Sigma\ \sigma_m(2,5) - 5.78*\Sigma\ \sigma_p(3,4) \qquad (9)$$
$$n=25 \quad r^2=0.965 \quad s=0.477 \quad F=306.8$$

Figure 3 shows a scatter plot of predicted vs measured pKa for Equation 8. The σ_m values for the (2,5) positions, the σ_p values for the (3,4) positions contribute the most weight to the equation. Though we examined other possibilities, we found that using σ_m values at the (2,5) positions and σ_p values at the (3,4) positions led to the most accurate equations.

A Graphical Analysis of the Pyrrole Data. The QSAR analysis of the pyrrole insecticides began with a graphical analysis of the data. The graphs were used to detect obvious trends and to aid in deciding which parameters were needed to develop a quantitative model. The coverage of the data space is shown in Figure 4. A broad range in both LogP and pK_a dimensions are covered though a point or two in the high LogP, low pK_a quadrant would be useful.

The relationship of the biological activity to the descriptors is shown in the following figures. Figure 5 shows the relationship between LogP and the biological response. The activity reaches a maximum at a LogP of about 5.5 to 6.0 and then seems to level off. The scatter plot indicates that either a linear or perhaps a quadratic or bilinear model will be needed to describe the LogP dimension.

Figure 6 shows the relationship between activity and pK_a. Clearly the activity is much more sensitive to changes in the pK_a of the compounds. The response drops off steeply on either side of the optimum pK_a of approximately 7. Also note that the

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

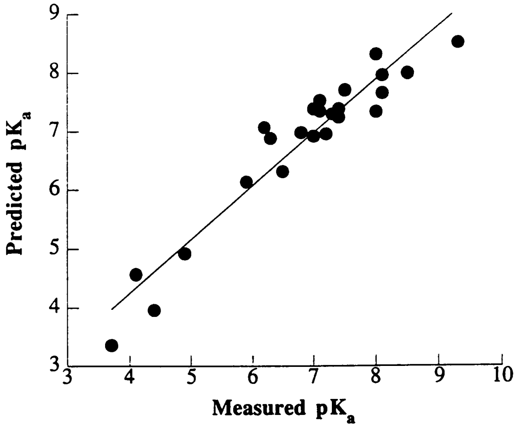


Figure 3. Predicted vs measured pKa based upon Equation 8.

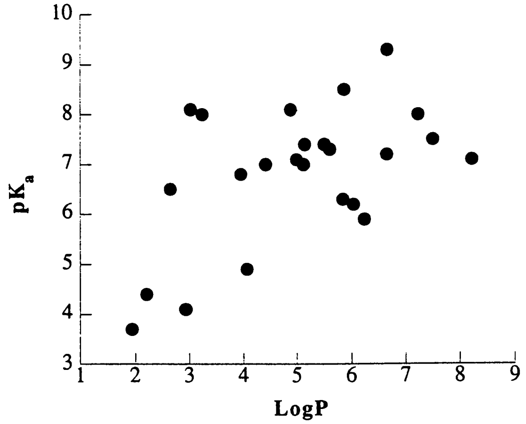


Figure 4. Plot of pKa vs LogP for the pyrrole insecticides.

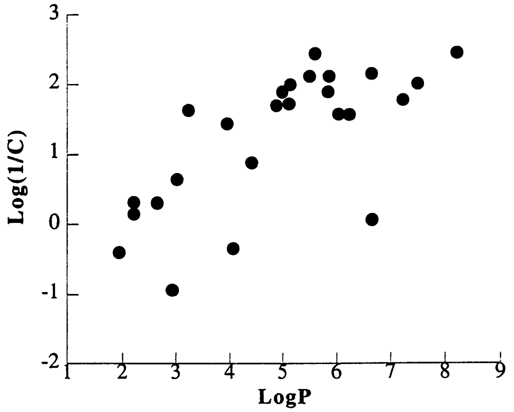


Figure 5. Plot of biological activity vs LogP for the pyrrole insecticides.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

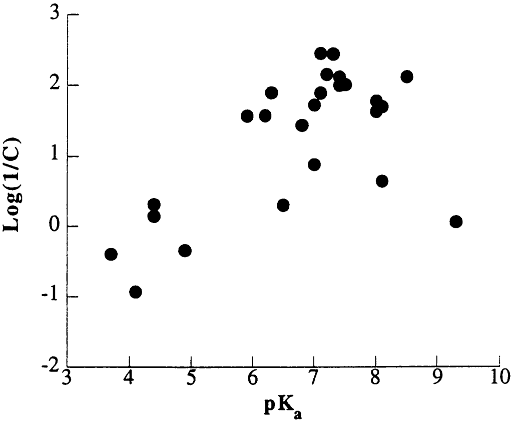


Figure 6. Plot of biological activity vs pKa for the pyrrole insecticides.

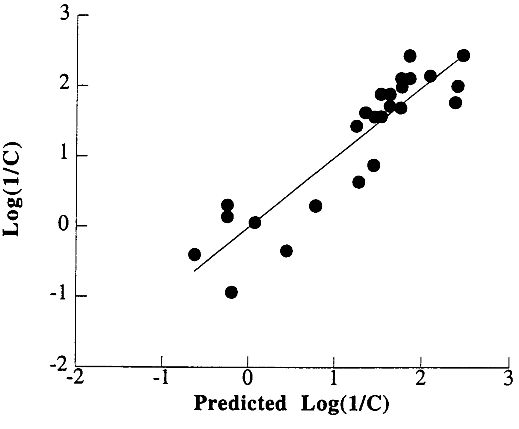


Figure 7. Plot of activity vs predicted activity based upon Equation 10.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

response is not symmetric about the maximum; the response drops off more sharply at high pK_a's than low pK_a's. A quadratic model will not work well here because of the lack of symmetry about the maximum. A bilinear model seems required.

The combined information from Figures 5 and 6 indicate that the points lie on a three dimensional hill or ridge in parameter space. With the aid of computer graphics the shape of the dataset can be confirmed by examination of a three dimensional scatter plot .

The QSAR of the Pyrrole Insecticides.

Based upon the graphical analysis of the data we decided to develop and examine two models to describe the observed biological response. In both models the pK_a and LogP are the important factors. The first model is linear in LogP and bilinear in pK_a (Equation 10) (*26-28*). The r^2 is 0.81, indicating that 81% of the variation in the response is explained by the model.

$$\mathrm{Log}(1/C) = 0.258{*}\mathrm{LogP} + 0.441{*}pK_a - 6.26{*}\mathrm{Log}(\mathit{B}K_a + 1) - 2.76 \qquad (10)$$

$$n=25 \quad s=0.442 \quad r^2=0.82 \quad F=32.53 \quad \mathrm{Log}(\mathit{B})= -9.0$$

An examination of the plot of predicted activity vs measured activity shows a pattern (Figure 7). Activity is consistently over predicted at high levels, under predicted for somewhat less active compounds, and over predicted for compounds in the middle of the activity range. The pattern is more clearly shown in the residual plot (Figure 8). The pattern present in the plot of residuals vs LogP indicates that the model is failing to accurately describe the relationship of LogP to activity. No such pattern is present in a plot of residuals vs pK_a. The problem appears to be caused by the use of a linear term to describe a non-linear LogP relationship.

In the Equation 11 we address the shortcomings of Equation 10 by introducing a bilinear term for LogP. In Equation 11 both the LogP and the pK_a are modeled using the bilinear functional form. The r^2 and standard deviation both improve in Equation 11 while F worsens somewhat with the inclusion of an additional term in the equation.

$$\begin{aligned}\mathrm{Log}(1/C) = {} & 0.348{*}\mathrm{LogP} - 0.312{*}\mathrm{Log}(\mathit{A}\mathrm{P} + 1)\\ & + 0.409{*}pK_a - 6.13{*}\mathrm{Log}(\mathit{B}K_a + 1) - 2.89 \qquad (11)\end{aligned}$$

$$n=25 \quad s=0.434 \quad r^2=0.83 \quad F=25.26 \quad \mathrm{Log}(\mathit{A})= -6.0 \quad \mathrm{Log}(\mathit{B})= -9.0$$

An examination of Figure 9, the plot of measured vs predicted activity for Equation 11 shows that the systematic pattern present in Figure 7 has been eliminated. The residual plot shown in Figure 10 confirms the fact that the systematic bias has been removed. Unfortunately the bilinear term in Equation 11 is only significant at the 80% level, therefore there is a 20% possibility of a chance correlation. However, we feel that inclusion of the term is justified for two reasons. As shown by the residual plot, Equation 11 provides a better fit to the data because the model lacks the systematic bias present in Equation 10. Also, in other studies conducted on more extensive pyrrole datasets the bilinear term is highly significant.

Equation 11 describes a ridge in parameter space. The response surface generated by the model is shown in Figure 11. The surface is a sloping ridge. The optimum LogP is about 5.5 and the optimum pK_a is about 7.2. These optima agree well with the information previously published on other classes of uncouplers. The pK_a optimum in our system is probably shifted above a pK_a of 7.0 due to problems involved in extrapolating the measurements made in a mixed solvent system to generate a pK_a value for pure water. Equation 11 has been used successfully to predict the

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

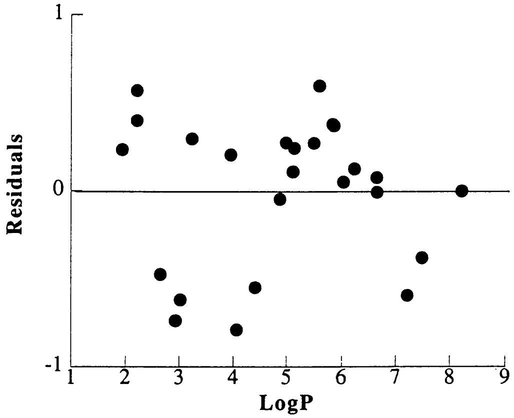


Figure 8. Plot of residuals vs LogP based upon Equation 10.

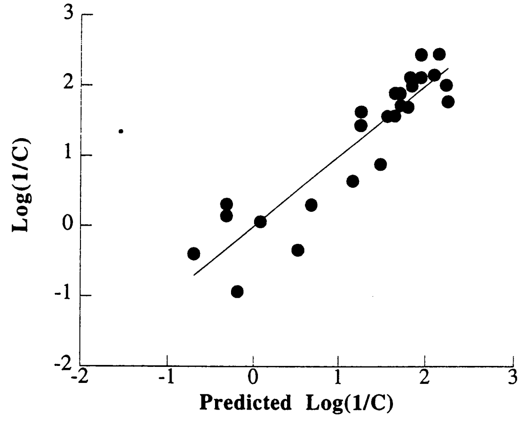


Figure 9. Plot of activity vs predicted activity based upon Equation 11.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

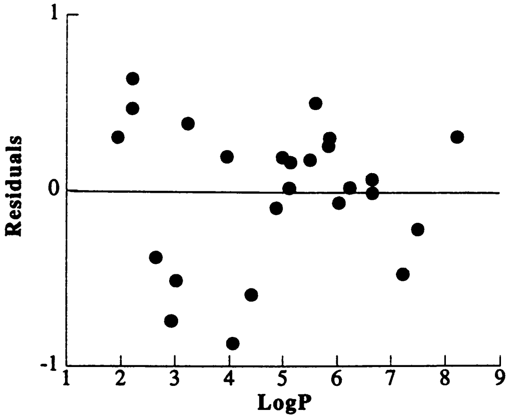


Figure 10. Plot of residuals vs LogP based upon Equation 11.

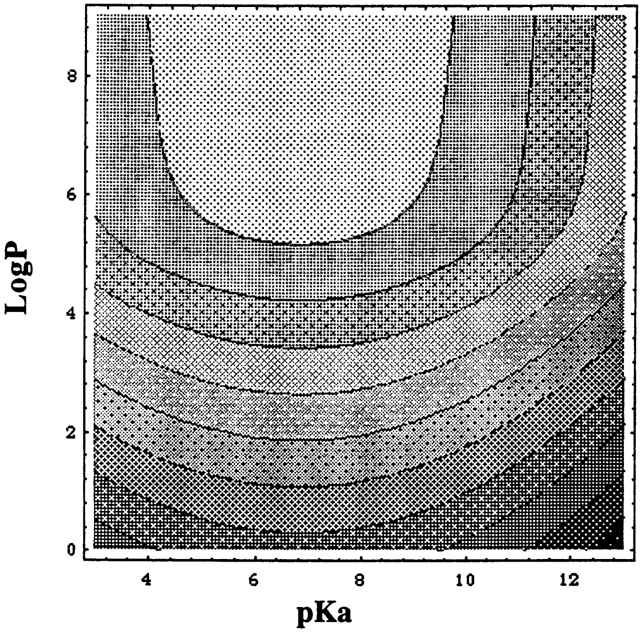


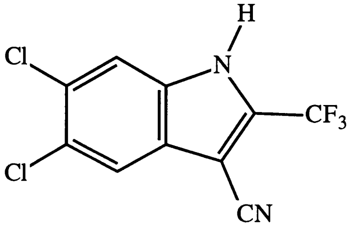
Figure 11. Contour plot of activity vs LogP and pK_a based upon Equation 11.

activity of new pyrroles. It has allowed us to select functional groups that are compatible with maintaining or increasing the biological response.

Comparison of our models with models previously published, and comparison of our structures with the structures of known uncouplers reaffirmed the conclusion that uncoupling is solely a function of physical properties. If the LogP and pK_a are in the right range, and if the compounds are metabolically stable they should act as uncouplers.

The Design of Indole Uncouplers

As we studied the QSAR of the pyrrole insecticides we began to look for other heterocycles that might also show uncoupling activity. Judging from published studies on uncoupling and our own experience with the pyrroles it seemed that any heterocycle with an acidic hydrogen that could be made sufficiently lipophilic should be an uncoupler. The indole ring system suggested itself for a number of reasons. First, both triazole and benzotriazole systems are known to act as uncouplers. Therefore since pyrroles could act as uncouplers it seemed likely that benzopyrroles or indoles, could also have the same activity. Quick back-of-the-envelope calculations showed that the indole NH proton could be properly acidic given appropriate substituents on the indole ring. When we examined the results of CLOGP predictions for candidate indoles the results were also encouraging. Finally, an examination of the patent literature indicated that the indoles we were considering had not been patented. A synthesis effort was begun and our prediction was proven correct. Indole uncouplers are highly active against Southern Armyworm. The synthetic effort is continuing and the first patent application has been submitted.



Conclusions

We have shown that the primary, and perhaps the exclusive, mode of action of the pyrrole insecticides is the uncoupling of oxidative phosphorylation. In common with other classes of uncouplers the activity of the pyrroles is a function of the physical properties of the compounds. In addition to developing an accurate model for the prediction of pyrrole pK_a's we have developed QSAR models that describe as much as 83% of the variation of the biological response. Finally we have successfully used our pyrrole QSAR model to predict the activity of a new class of uncouplers based upon substituted indoles. Our prediction was confirmed by the subsequent synthesis of highly active indole insecticides with uncoupling activity.

References

1. Carter, G. T.; Nietsche, J. A. ; Goodman, J. J.; Torrey, M. J.; Dunne, T. S.; Borders, D. B.; Testa, R. T. *J. Antibiotic.* **1987**, *40*, 233.
2. Cazeneuve, P. and Lepine, R. *Compt. Rend. Acad. Sci.* **1885**, *100,* 1167.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

3. Spencer, E.Y. *Guide to the Chemicals used in Crop Protection, 6th ed.* **1973,** Res. Branch Agric. Can. Ottawa, pp 225, 228, 243.
4. McLaughlin, S.G.A. and Dilger, J.P. *Physiological Reviews* **1980**, *60*, 825.
5. Astwood, E.B. *The Pharmacological Basis of Therapeutics 4th ed.* Goodman, L.S. and Gilman, A. (eds.) **1970**, MacMillan: New York, 1481.
6. Loomis, W.F., and Lipmann, F. *J. Biol. Chem.* **1948,** *173*, 807.
7. Mitchell, P. *Nature* **1961**, *191*, 144.
8. Mitchell, P. *Biol. Rev.* **1966,** *41*, 445.
9. Hansch, C.; Kiehs, K.; and Lawrence, G. *J. Am. Chem. Soc.* **1965**, *87*, 5770.
10. Tollenaere, J. *J.Med.Chem.*, **1973**, *16*, 791.
11. Miyoshi, H. and Fujita, T. *Biochim. Biophys. Acta* **1988**, *935*, 312.
12. Jones, O. and Watson, W. *Biochem. J.* **1967**, *102*, 564.
13. Uda, M.; Toyooka, K.; Horie, K.; Shibuya, M.; Kubota, S.; and Terada, H. *J.Med.Chem.* **1982**, *25*, 557.
14. Büchel, K. and Draber, K. *Adv. in Chem. Ser.* **1972**, *114*, 141.
15. Miyoshi, H.; Nishioka, T.; and Fujita, T. *Biochim. Biophys. Acta,* **1987**, *891*, 293.
16. Miyoshi, H.; Nishioka, T.; and Fujita, T. *Biochim. Biophys. Acta* , **1987**, *891*, 194.
17. Ravanel, P.; Tissut, M.; and Douce, R. *Phytochemistry* , **1982**, *21*, 2845.
18. Balaz, S.; Sturdik, E.; Durcova, E.; Antalik, M.; and Sulo, P. *Biochim. Biophys. Acta* **1986**, *851*, 93.
19. Terada, H. *Biochim. Biophys. Acta* **1981**, *639*, 225.
20. Labbe-Bois, R.; Laruelle, C.; and Godfroid, J. *J.Med.Chem.* **1975**, *18*, 85.
21. Cajina-Quezada, M. and Schultz, T. W. *Aquatic Tox.* **1990**, *17*, 239.
22. Heytler, P. *Pharmacol. Ther.* **1980**, *10*, 461.
23. Chou, J. and Jurs, P. *J. Chem. Inform. Comput. Sci.* **1979**, *19,* 172.
24. Hansch, C. and Leo, A. *Exploring QSAR;* American Chemical Society: Washington, D.C., 1995; Vol. 1, 125-168.
25. Barlin, G. B. and Perrin, D.D. *Quarterly Reviews*, **1966**, *20*, 75-101.
26. Kubinyi, H. *Arneim.-Forsh. / Drug Res.* **1976,** *26*, 1991.
27. Kubinyi, H. *J.Med.Chem.* **1977**, *20*, 625.
28. Kubinyi, H. and Kehrhahn, O.-H. *Arneim.-Forsh. / Drug Res.* **1976**, *28*, 598.

RECEIVED May 25, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch015

Chapter 16

Quantitative Structure–Activity and Molecular Modeling Studies of Novel Fungicides and Herbicides Having 1,2,4-Thiadiazoline Structures

Akira Nakayama, Kenji Hagiwara, Sho Hashimoto, and Hideo Hosaka

**Odawara Research Center, Nippon Soda Company, Ltd., Takada,
Odawara 250–02, Japan**

Structure-activity analyses of bioactive 1,2,4-thiadiazolines were studied. With the procedure of the traditional quantitative structure-activity analyses, the fungicidal activity of a series of Δ^3-1,2,4-thiadiazoline derivatives against the cucumber downy mildew was examined using a parameter for the reactivity of molecules defined by quantum-chemical indices. The result suggested that a suitable range of the reactivity is required for this series of compounds to control the plant disease through the inhibition of SH-enzymes. A series of the Δ^2-1,2,4-thiadiazolines exhibited a light-dependent herbicidal activity similar to that induced by cyclic imide herbicides. The structure-activity analyses of the light-dependent compounds were made by using theoretically estimated steric and electrostatic features of the molecule. Similarity indices among these compounds rationalized their structure-activity patterns stereoelectronically. The structural requirements for light-dependent herbicides were hypothesized on the basis of the molecular similarity indices.

1,2,4-Thiadiazolines have been attractive candidates of new drugs and agrochemicals since their biological activities have scarcely been reported. There are three isomeric forms, *i.e.*, Δ^2, Δ^3 and Δ^4, in the 1,2,4-thiadiazolines according to the position of the double bond as shown in Figure 1. The syntheses of derivatives have also been interesting subjects in synthetic chemistry. We have synthesized a number of derivatives of these isomeric thiadiazolines and screened them for their biological activities to develop new pesticides. The derivatives have been shown to exhibit various pesticidal activities. For example, a series of Δ^2-1,2,4-thiadiazolines having a fused ring system (**I**) show a herbicidal activity (*1*), and the 5-phenylimino-Δ^3-1,2,4-thiadiazolines (**II**) exhibit a fungicidal activity to control a plant diseases (*2*). Furthermore, the Δ^4-1,2,4-thiadiazoline derivatives (**III**) are insecticidal (*3*). This chapter describes the structure-activity studies of the Δ^3- and Δ^2-1,2,4-thiadiazolines based on the traditional Hansch-type quantitative structure-activity analyses (QSAR) and molecular modeling studies, respectively.

0097–6156/95/0606–0213$12.00/0

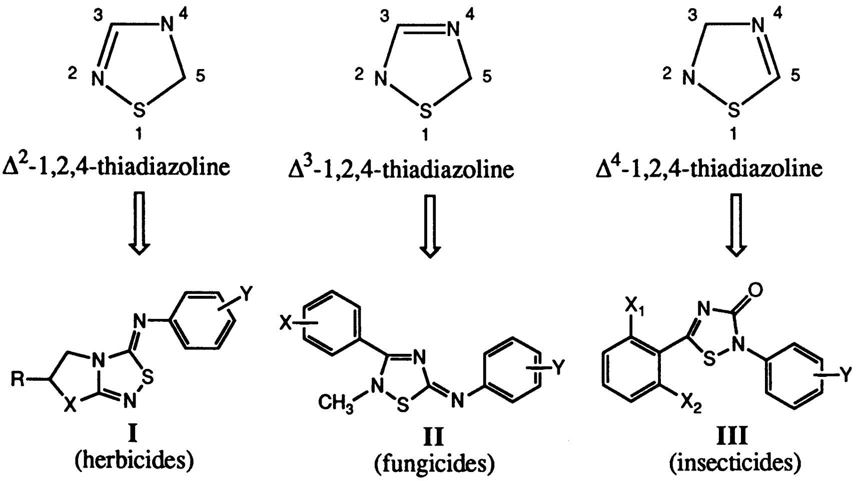
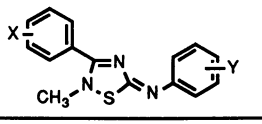


Figure 1. Isomers of the 1,2,4-thiadiazoline and their derivatives showing pesticidal activity.

Structure-Activity Analyses of Fungicidal Δ^3-1,2,4-Thiadiazolines

In the screening studies of the Δ^3-1,2,4-thiadiazolines, a series of compounds having 3-phenyl and 5-phenylimino groups (**II**) have been found to show various degrees of a fungicidal activity against the cucumber downy mildew (*Pseudoperonospora cubensis*) (*2*). Table I shows the preventive activity of some representative compounds which are effective to control the disease at lower than 10 ppm in the pot test. As shown below, the mode-of-action and chemical reactivity studies have suggested that a specific type of the reactivity is related to the antifungal activity. Thus, the reactivity-activity correlation was analyzed by the traditional QSAR procedure.

Table I. Preventive Activity of the Δ^3-1,2,4-Thiadiazolines against the Cucumber Downy Mildew (*P. cubensis*)[a]



No.	X	Y	EC_{50} (ppm)[b]
1	H	4-Cl	5.7
2	4-Cl	4-Cl	7.2
3	4-Cl	4-Me	9.2

[a] Adapted from ref. 2.

[b] Assessed by the pot test.

Mode of Action: The fungicidal mode of action was investigated against the cucumber downy mildew using the most active compound (**1**). The results can be summarized as follows;

1. Compound **1** exhibited a potent preventive effect against the disease, but it was neither curative nor systemic.
2. The inhibition of the spore germination and the burst of the zoospores of the fungus were observed, but the addition of cysteine canceled these effects on the zoospores.
3. Compound **1** inhibited the activity of the alcohol dehydrogenase.

These fungitoxic and enzyme inhibiting properties were similar to those of the SH-inhibiting fungicides such as captan and chlorthalonil (*4, 5*). The results strongly suggested that the primary mode of action of compound **1** is the inhibition of the SH-enzyme (*2*).

Chemical Reactivity: The fungicidal 3-phenyl-5-phenylimino-Δ^3-1,2,4-thiadiazolines (**II**) were found to react *in vitro* with SH-compounds such as mercaptans. Compound **1**, for example, reacted with benzyl mercaptan under mild conditions to give a thiourea derivative as shown in Figure 2. This reaction looks like a model of the biological reaction with SH-enzymes. The chemical reactivity and the fungitoxic activity of compound **1** and its analogs were examined and listed in Table II. The analogs having the 1,2,4-oxadiazoline and 1,2,4-thiadiazole structures, compounds **4** and **5**, neither reacted with benzyl mercaptan nor showed fungitoxic activity. Compound **6**, the 5-benzoylimino analog of **1**, which presumably exists in the form of a heteropentalene as the more stable tautomer (**6'**), was also much less active than compound **1**, and did not react with the mercaptan. These results strongly suggested that the reactivity with the SH group is of critical importance for the fungicidal activity (*2*).

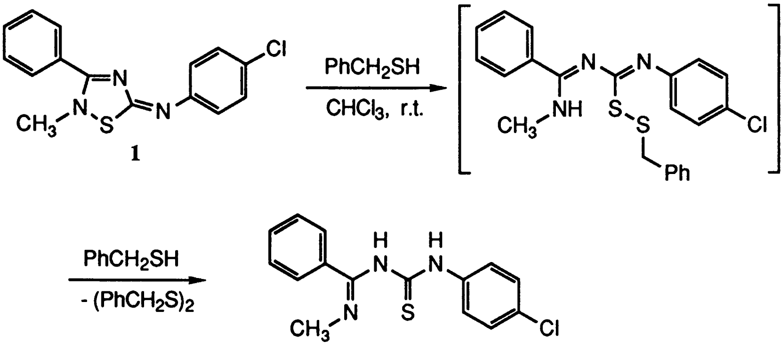


Figure 2. Reaction of a fungicidal Δ^3-1,2,4-thiadiazoline (**1**) with benzyl mercaptan.

Molecular Orbital Parameters: To understand the role of chemical reactivity by the structure-activity analyses, the introduction of the semi-empirical molecular orbital parameters was examined. We employed the MNDO-PM3 procedure (*6, 7*) in the MOPAC program (QCPE #455, version 6.0), and calculated various quantum-chemical indices of active and inactive compounds. We found that the distribution of the HOMO electrons is significantly different between active and inactive molecules as shown in Figure 3. In the active molecule (**1**), the HOMO is localized mainly on the sulfur of the 1,2,4-thiadiazoline ring, whereas the HOMO density at the corresponding position in the inactive molecules (**4**, **5** and **6'**) is much lower. The reactivity of the electrons of HOMO at the sulfur of compound **1** and that at the corresponding position in related compounds were considered to be related to the biological activity (*8*).

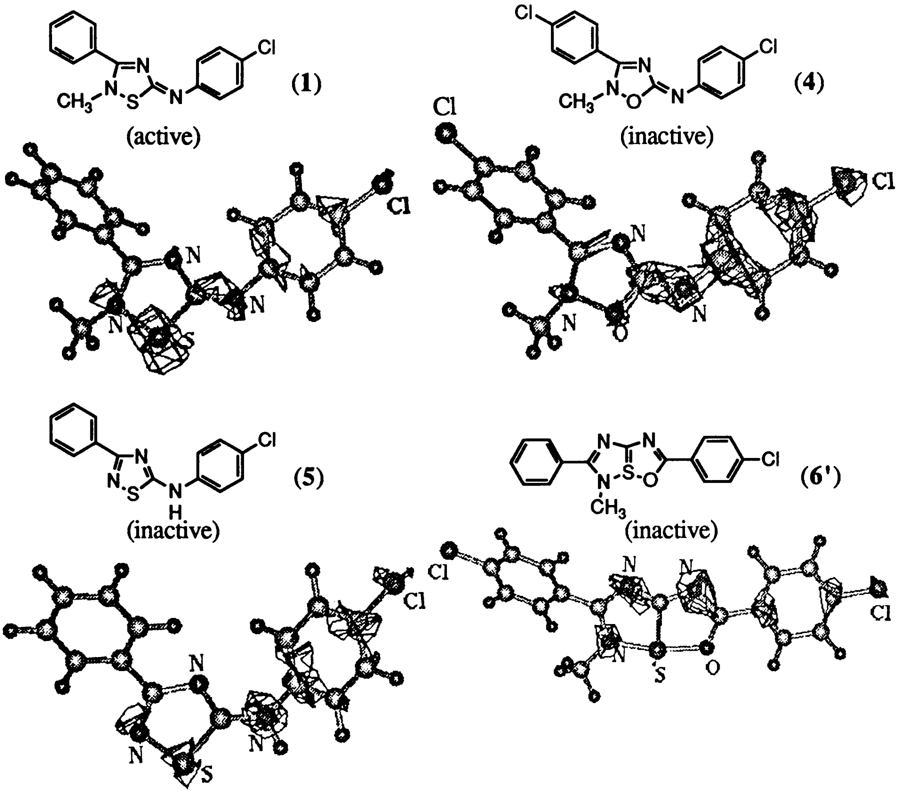


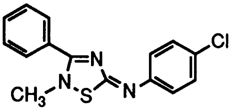
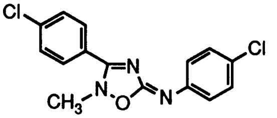
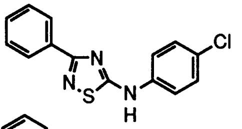
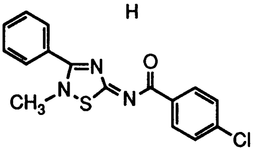
Figure 3. HOMO distributions in active and inactive Δ^3-1,2,4-thiadiazoline and related compounds (Reproduced with permission from ref. 8. Copyright 1993, VCH Verlag).

We have defined a reactivity index expressed as equation 1 as a quantitative measure of the reactivity.

$$R(i) = f_r(i) / -E_{HOMO} \times 10^2 \qquad (1)$$

In equation 1, $f_r(i)$ is the frontier electron density of the HOMO at the atom i, and $-E_{HOMO}$ is the unsigned energy of HOMO in eV measured from the zero level. The $f_r(i)$ represents the localizability of the HOMO electrons at atom i within a molecule, and $-E_{HOMO}$ expresses the relative electron stability among molecules. With this definition, the lower the $-E_{HOMO}$ value, *i.e.*, the less stable the frontier electrons, as well as the higher the $f_r(i)$ value, the higher the relative reactivity index, R(i), among molecules compared (*9*). The R(i) indices for the sulfur at the 1-position of thiadiazoline and thiadiazole derivatives (**1**, **5** and **6'**) and that for the 1-oxygen of the oxadiazoline analog (**4**) were calculated, the values being listed in Table II. The R(1) index of the active compound **1** was 6.47, whereas the values for inactive compounds **4** - **6'** were much lower (*8*). Thus, the rate determining step of the reaction shown in Figure 2 should include a nucleophilic attack of the sulfur at the 1-position of compound **1** against the SH sulfur of the mercaptan. The R(1) index is , in fact, similar to the superdelocalizability to compare the frontier electron reactivities at a corresponding position among congeners (*10*).

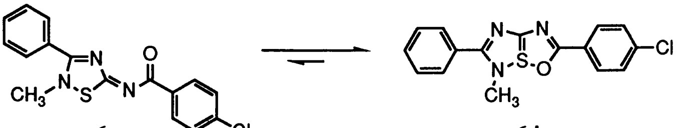
Table II. Biological Activities, R(1) Index and Reactivity of Antifungally Active and Inactive Compounds[a]

Compound (No.)		EC_{50} (*P. cubensis*) (ppm)		R(1)	Reactivity with benzyl mercaptan
		in vivo [b]	*in vitro* [c]		
	(1)	5.7	0.93	6.47	reacted
	(4)	400	>100	1.07	not reacted
	(5)	>400	>100	2.21	not reacted
	(6)	>400	*ca.* 50	1.23[d]	not reacted

[a] Reproduced with permission from ref. 8. Copyright 1993, VCH Verlag.

[b] Assessed by the pot test. [c] Inhibition of the spore germination.

[d] Calculated for the tautomeric form **6'**.



6 **6'**

Quantitative Structure-Activity Relationships (*8*): The reactivity index R(1) was calculated for the 2-methyl-3-phenyl-5-phenylimino-Δ^3-1,2,4-thiadiazolines (**II**) having various substituents on the two benzene rings, as shown in Table III, and the correlation with the fungicidal activity was examined. For the fungicidal activity index, pEC_{50} (= – $\log EC_{50}$) against the cucumber downy mildew was used. The EC_{50}

is the molar concentration required to inhibit the infection by 50% of the control assessed by the pot test.

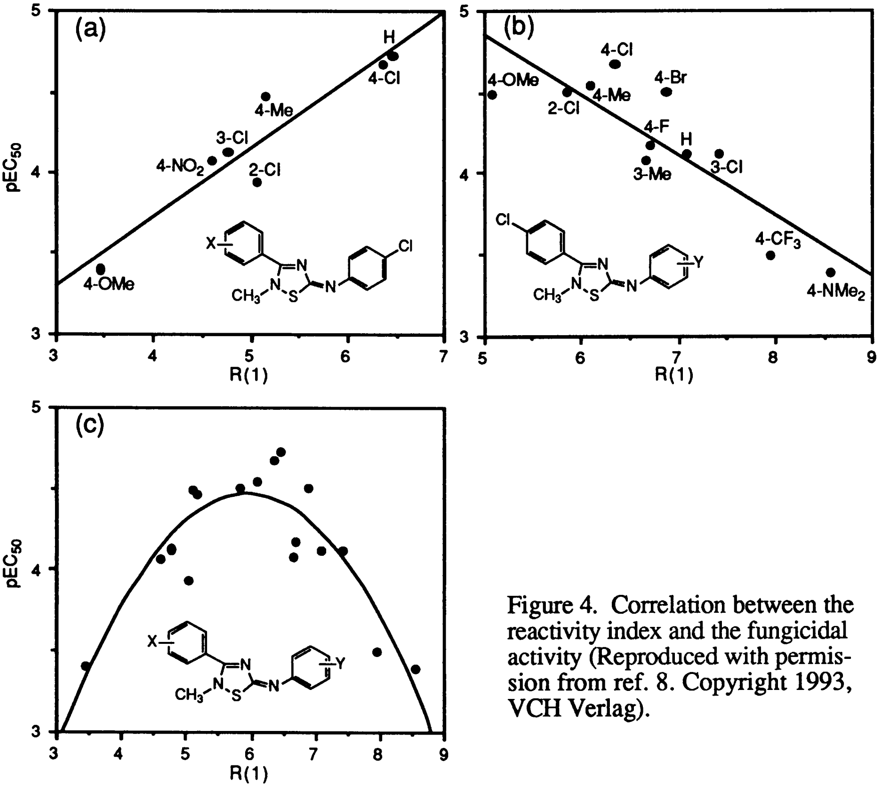


Figure 4. Correlation between the reactivity index and the fungicidal activity (Reproduced with permission from ref. 8. Copyright 1993, VCH Verlag).

For compounds in which the single substituent (X) on the 3-phenyl moiety is varied, but the substituent (Y) on the 5-phenylimino moiety is fixed as the 4-Cl (compounds 1, 2, **7** - **11**), the R(1) value is in a range of *ca.* 3.5 - 6.4, and a positive correlation between pEC_{50} and R(1) was observed as shown by equation 2 and Figure 4(a).

$$pEC_{50} = 0.42(\pm 0.17)R(1) + 2.04 \qquad (2)$$

$$n = 7,\ r = 0.94,\ s = 0.17,\ F_{1,5} = 39.46$$

In the above and subsequent equations, n represents the number of compounds, r the multiple correlation coefficient, s the standard deviation and F the ratio of regression and residual variances. The figures in parentheses are the 95% confidence interval.

For compounds (**3**, **8**, **13** - **21**) in which the single substituent (Y) is varied, but the substituent (X) is fixed as the 4-Cl, the R(1) index is distributed in a range of *ca.* 5.1 – 8.6, and the correlation between pEC_{50} and R(1) was negative as shown by equation 3 and Figure 4(b).

$$pEC_{50} = -0.37(\pm 0.17)R(1) + 6.70 \qquad (3)$$

$$n = 11, \; r = 0.85, \; s = 0.23, \; F_{1,9} = 23.89$$

For the combined data set of compounds used for equations 2 and 3, equation 4 was formulated showing a parabolic relationship between pEC_{50} and R(1) as illustrated in Figure 4(c).

$$pEC_{50} = 2.14(\pm 0.67)R(1) - 0.18(\pm 0.06)R(1)^2 - 1.91 \qquad (4)$$

$$n = 17, \; r = 0.88, \; s = 0.21, \; F_{2,14} = 24.76$$

Table III. QSAR Parameters and pEC_{50} Value of Fungicidal Δ^3-1,2,4-Thiadiazolines (II)[a]

Compd.	Substituents		Parameters			pEC_{50}	
No.	X	Y	R(1)	π_X	π_Y	obsd.	calcd.[b]
1	H	4-Cl	6.47	0.00	0.93	4.72	4.48
2	2-Cl	4-Cl	5.16	-0.05	0.93	4.46	4.40
3	4-Cl	4-Me	6.10	0.93	0.46	4.54	4.43
7	3-Cl	4-Cl	4.77	0.99	0.93	4.12	4.07
8	4-Cl	4-Cl	6.36	0.93	0.93	4.67	4.33
9	4-Me	4-Cl	5.06	0.54	0.93	3.93	4.26
10	4-NO_2	4-Cl	4.61	0.16	0.93	4.07	4.15
11	4-OMe	4-Cl	3.46	0.06	0.93	3.40	3.40
12	3,5-Cl_2	4-Cl	5.11	1.98	0.93	3.74	4.10
13	4-Cl	H	7.08	0.93	0.00	4.11	4.32
14	4-Cl	2-Cl	5.86	0.93	1.00	4.50	4.33
15	4-Cl	3-Cl	7.43	0.93	0.98	4.11	4.00
16	4-Cl	3-Me	6.66	0.93	0.50	4.08	4.35
17	4-Cl	4-F	6.71	0.93	0.25	4.16	4.39
18	4-Cl	4-Br	6.88	0.93	1.36	4.50	4.14
19	4-Cl	4-CF_3	7.96	0.93	1.05	3.49	3.68
20	4-Cl	4-OMe	5.10	0.93	0.05	4.48	4.36
21	4-Cl	4-NMe_2	8.57[c]	0.93	0.68	3.39	3.29
22	4-Cl	4-Et	6.24	0.93	0.95	4.18	4.33
23	4-Cl	4-Pr	5.90	0.93	1.48	3.16	4.24
24	4-Cl	4-iPr	6.04	0.93	1.46	4.39	4.25
25	4-Cl	4-Bu	6.00	0.93	2.06	4.07	4.14
26	4-Cl	3,4-Cl_2	6.13	0.93	1.79	4.18	4.19
27	4-Cl	2,3-Me_2	4.54	0.93	1.88	4.12	3.98
28	4-Cl	3,4-Me2	5.22	0.93	0.99	4.40	4.23
29	4-Cl	3,5-Me_2	6.36	0.93	1.00	4.10	4.31

[a] Reproduced with permission from ref. 8. Copyright 1993, VCH Verlag.

[b] Calculated by equation 5. [c] Calculated for the electron density on the next-HOMO.

The QSAR was subsequently examined for a larger set of compounds having the higher alkyl groups and the disubstitutions on either of the rings (Compounds **1** - **3**, **7**

- **29**). In this case, the addition of the hydrophobicity term ($\Sigma\pi$) was significant as well as the quadratic terms of the reactivity index affording equation 5.

$$pEC_{50} = 2.04(\pm 0.61)R(1) - 0.17(\pm 0.05)R(1)^2 - 0.18(\pm 0.15)\Sigma\pi - 1.45 \quad (5)$$

$$n = 26,\ r = 0.84,\ s = 0.21,\ F_{3,22} = 17.04$$

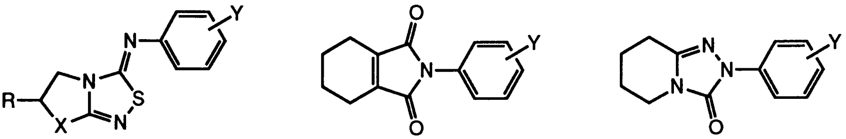
In equation 5, $\Sigma\pi$ is the summation of π values for X and Y substituents in structure **II**. The value of π_X was approximated by that from log P values of substituted benzoylhydrazines, and the π_Y value by the π value from substituted anilines. The reference log P values were taken from the log P database from the Pomona College Medicinal Chemistry Project and other reference (*11*). The negative coefficient of the $\Sigma\pi$ term suggests that the lower hydrophobicity is more favorable for the activity. The parabolic relationship between pEC_{50} and R(1) predicts the presence of an optimum R(1) which is *ca.* 6.0. No Hammett-type σ constants was able to replace the R(1) parameter. This is probably due to perturbations of the "regular" electronic effect of substituents on the 1-sulfur atom by complex heteroatomic environments. The values of parameters used in the analyses and pEC_{50} values observed and calculated by equation 5 are listed in Table III.

The reactivity index, R(1), was a useful parameter to account for the activity in above correlation equations. The SH-inhibiting fungicides are known to exhibit their activity by reacting with SH-groups in fungal cells (*12, 13*). However, many SH-inhibitors such as captan and folpet are believed to react nonselectively with proteins or other cellular components having SH, OH, and NH_2 groups (*14*). The presence of the optimum value in the reactivity index suggests that a molecule with excessive reactivity may be trapped by non-target components or decomposed by unexpected reactions during the process of transport to the target sites. A suitable range of reactivity should be required to control the plant disease through the inhibition of SH-enzymes (*8*).

Structure-Activity Analyses of Herbicidal Δ^2-1,2,4-Thiadiazolines and Light-dependent Herbicides

During the screening process of the Δ^2-1,2,4-thiadiazoline derivatives, a class of compounds having the 5-phenylimino group and a fused ring structure (**I**) was found to show a potent herbicidal activity (*1*). As described below, their phytotoxic symptoms and effects of benzene ring substituents on the activity were similar to those observed for cyclic imide and amide herbicides such as tetrahydrophthalimides (**IV**) and 1,2,4-triazolin-3-ones (**V**) (Figure 5). At the first glance, however, their chemical structures are not similar to each other, differing in the structure of the heterocyclic ring system as well as the position of the "anilino" nitrogen. The "anilino" nitrogen in the cyclic imide and amide herbicides (**IV** and **V**) is endocyclic, whereas the corresponding nitrogen in the Δ^2-1,2,4-thiadiazolines is exocyclic. We have been interested in the way how these "dissimilar" herbicides are recognized by their receptor on the molecular level, and examined their possible similarity theoretically.

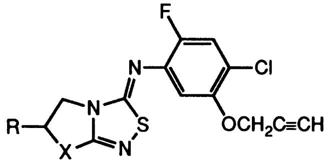
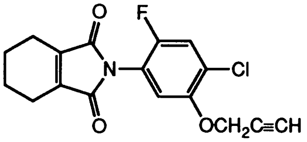
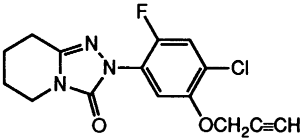
Herbicidal Activity and the Mode of Action: Herbicidal activity of Δ^2-1,2,4-thiadiazolines (**30** - **33**) against a redroot pigweed species (*Amaranthus retrofloxus*) is compared with that of corresponding tetrahydrophthalimide (**34**) and 1,2,4-triazolin-3-one (**35**) in Table IV.



I **IV** **V**

Figure 5. Structures of the herbicidal Δ^2-1,2,4-thiadiazolines (I) and the cyclic imide and amide herbicides (IV and V).

Table IV. Herbicidal Activity of Δ^2-1,2,4-Thiadiazolines (30–33) and Related Compounds (34, 35) against *A. retroflexus*

Structure	Compd. No.	ED_{90} (g/ha)[a]
	**30** ($X=CH_2$, R=Me)	*ca.* 8
	31 ($X=CH_2$, R=H)	31
	32 (X=S, $R=Me_2$)	63
	33 (X=O, R=Me)	63
	**34**	8
	**35**	8

[a] Post emergence activity in the green house.

In these compounds, the benzene-ring substitution pattern is fixed as the 2-fluoro-4-chloro-5-(2-propynyloxy). The herbicidal potency of the Δ^2-1,2,4-thiadiazoline (**30**) was *ca.* 8 g/ha which is almost equivalent to those of compounds **34** and **35** in terms of ED_{90}, the dose (g/ha) required to eradicate the weeds by 90% of the control (*15*). Compounds **30** - **33** as well as the imide and amide compounds **34** and **35** caused a bleaching symptom on treated leaves within 1-2 days after the application followed by wilting, desiccating and ultimate death. Such phytotoxic symptoms were only detected under the lighted conditions but not in the dark (*1*). The target of the phytotoxicity of the cyclic imide-type herbicides has been considered to be in the process of chlorophyll biosynthesis. Wakabayashi and Böger proposed that the inhibition of protoporphyrinogen oxidase with this type of herbicides induces an autooxidation of protoporphyrinogen affording an active oxygen species under the lighted conditions to peroxidize the lipid cell membranes (*16*).

The effect of the substituents (Y) of the benzene ring in the Δ^2-1,2,4-thiadiazolines (**I**) was examined using a number of compounds having various substitution patterns. The 2,4,5-tri-substitutions such as 2-fluoro-4-chloro-5-(2-

propynyloxy) were most effective to enhance the activity (*1*), being similar to substitution patterns for the cyclic imide herbicides (*17*).

These biological and chemical observations were taken to indicate that the Δ^2-1,2,4-thiadiazolines (I) probably share the target site or the mode of action in common with the cyclic imide and amide herbicides (IV and V) in spite of their dissimilar structures. One of the rationalizations of the "bioisosterism" in dissimilar compounds I, IV and V is to consider a possible rearrangement of Δ^2-1,2,4-thiadiazolines (I) to 1,2,4-triazoline-3-thiones (VI) in the biological systems. The structure of VI looks more similar to IV and V than the original structure (I) (Figure 6). We examined the stability of the Δ^2-1,2,4-thiadiazoline (32) under acidic, basic and UV-irradiated conditions. No isolated compound was identified, however, as that having the rearranged structure (VI). The original structure (I) was strongly suggested as the active form.

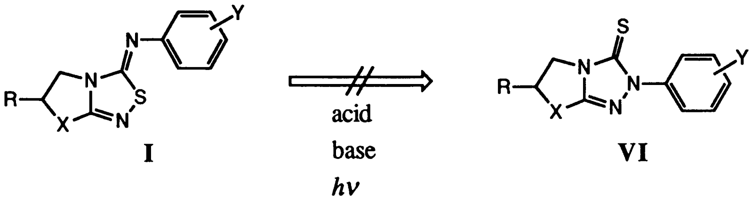


Figure 6. "Possible" rearrangement of the Δ^2-1,2,4-thiadiazolines (I) to the 1,2,4-triazoline-3-thiones (VI).

Molecular Electrostatic Potential and the Conformation: We have investigated the stereoelectronic features of the "bioisosteric" but dissimilar compounds theoretically selecting compounds **30**, **34** and **35**, in which the benzene-ring substituents are fixed to 2-fluoro-4-chloro-5-(2-propynyloxy). Semi-empirical molecular orbital calculations were performed to analyze their conformation and to estimate the molecular electrostatic potential field by the MNDO-PM3 method (*6*, *7*) in the MOPAC program.

The electrostatic property of the fused ring moiety of compounds **30**, **34** and **35** was examined using their simplified models (**30'**, **34'** and **35'**) in which the phenyl group in the original compounds was replaced by hydrogen. Their molecular electrostatic potential fields computed from their electrostatic-potential-derived atomic charges were drawn in Figure 7. The negative potential contours are distributed around the heterocyclic moieties, whereas the positive potential contours around the alicyclic moieties. The molecular electrostatic properties of their fused ring systems seem to be similar to each other in terms of the pattern in potential contour maps (*15*).

The conformation of the compounds **30**, **34** and **35** should mostly be governed by the potential energy of rotation around the bond connecting the fused heterocyclic ring with the benzene ring, because the ring systems are rigid. The energy variations with the rotation were calculated for the model structures (**30"**, **34"** and **35"**) in which the benzene-ring substitution pattern is the 2-fluoro-4-chloro. The energy diagrams shown in Figure 8 indicate that these molecules are flexible within some range of the rotation, considering an allowance of 0.5-1 kcal/mol from the minimum energy. They share a stable conformation in common when the torsional angle (θ) is *ca.* 240° as indicated in Figure 8d. Therefore, compounds **30**, **34** and **35** were considered also to take this conformation (θ=240°) at the receptor site in common. Their conformation was fixed in this way in the superimposition studies shown below (*15*).

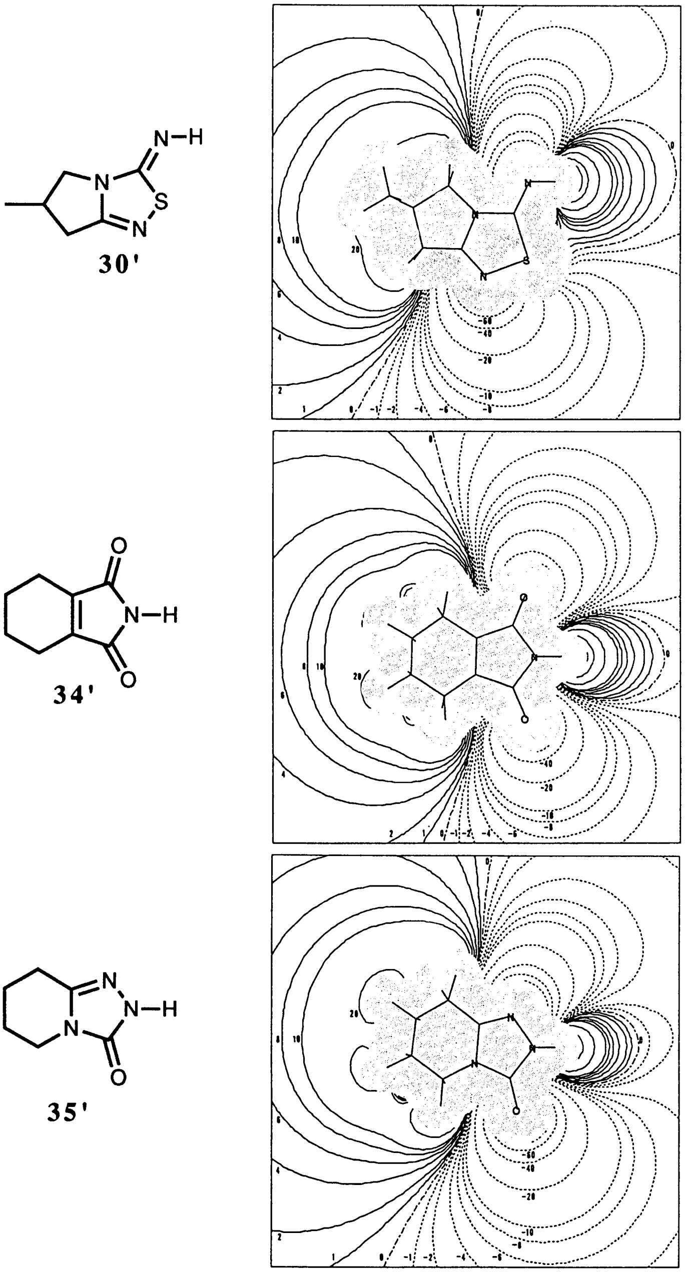


Figure 7. Molecular electrostatic potential contours of the fused ring systems in herbicidal compounds. Contour values are in unit of 10^{-3} au (Reproduced with permission from ref. 15. Copyright 1994, Pestic. Sci. Society of Japan).

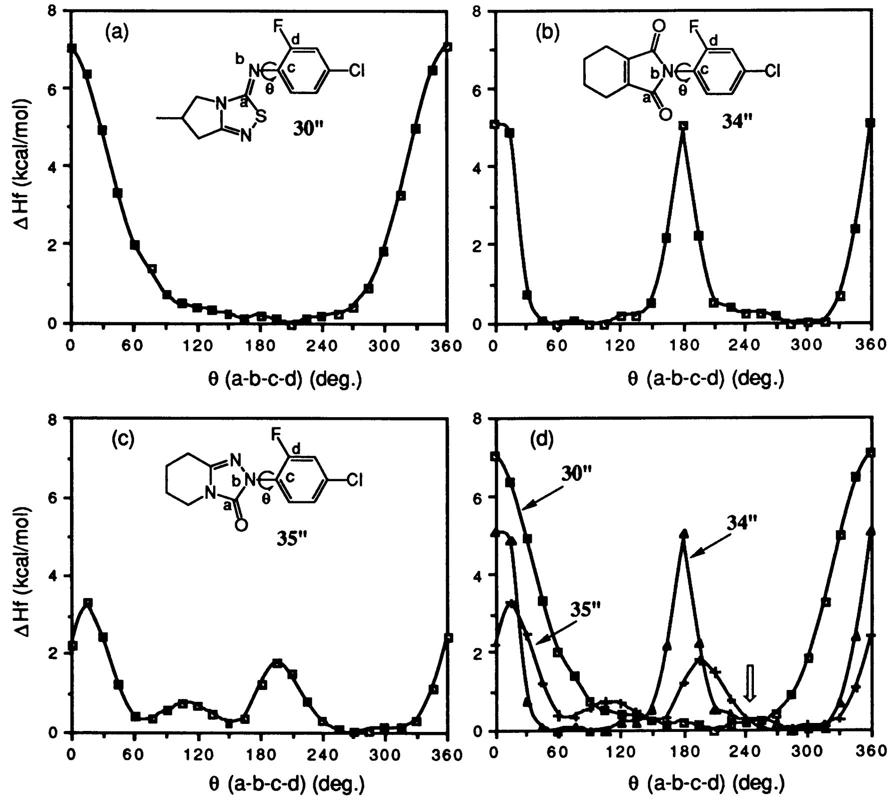


Figure 8. Potential energy diagrams for the rotation of the benzene ring in the model molecules . The torsional angle (θ) is defined as the dihedral angle determined by the positions of atoms a-b-c-d (Reproduced with permission from ref. 15. Copyright 1994, Pestic. Sci. Society of Japan).

Electrostatic Similarity: The electrostatic similarity of molecules was investigated by using the similarity index defined by Carbo *et al.* (*18*) and expanded by Richards & Hodgkin (*19, 20*) as shown in equation 6.

$$R_{AB} = \int \varepsilon_A \varepsilon_B d\tau \Big/ \left(\int \varepsilon_A^2 d\tau \int \varepsilon_B^2 d\tau \right)^{1/2} \qquad (6)$$

In equation 6, ε_A and ε_B are the electrostatic potentials on and outside the molecules A and B to be compared, respectively, and their values are computed from electrostatic-potential-derived atomic charges. The evaluation and an application of the electrostatic potential charges in QSAR studies had been reported previously (*21*). The index takes the value in the range of –1 to 1, with $R_{AB} = 1$ indicating the perfect similarity of species compared. The R_{AB} values for the two compounds among compounds **30, 34** and **35** were computed for pairs of structures superimposed by fitting only the benzene ring moieties each other using the least square fitting method. The values of ε_A and ε_B were calculated at each grid point of 0.3 Å intervals within the distance of 3.0 Å from the edge of van der Waals surface of each molecule. The space within the molecular surfaces of the two molecules were excluded from the calculation.

Table V. Electrostatic Similarity Index (R_{AB}) among Molecules 30, 34 and 35[a]

	30	34	35
30	1.000		
34	0.829[b] (0.841)[c]	1.000	
35	0.822[b] (0.831)[c]	0.926[b] (0.928)[c]	1.000

[a] Reproduced with permission from ref. 15. Copyright 1994, Pestic. Sci. Society of Japan

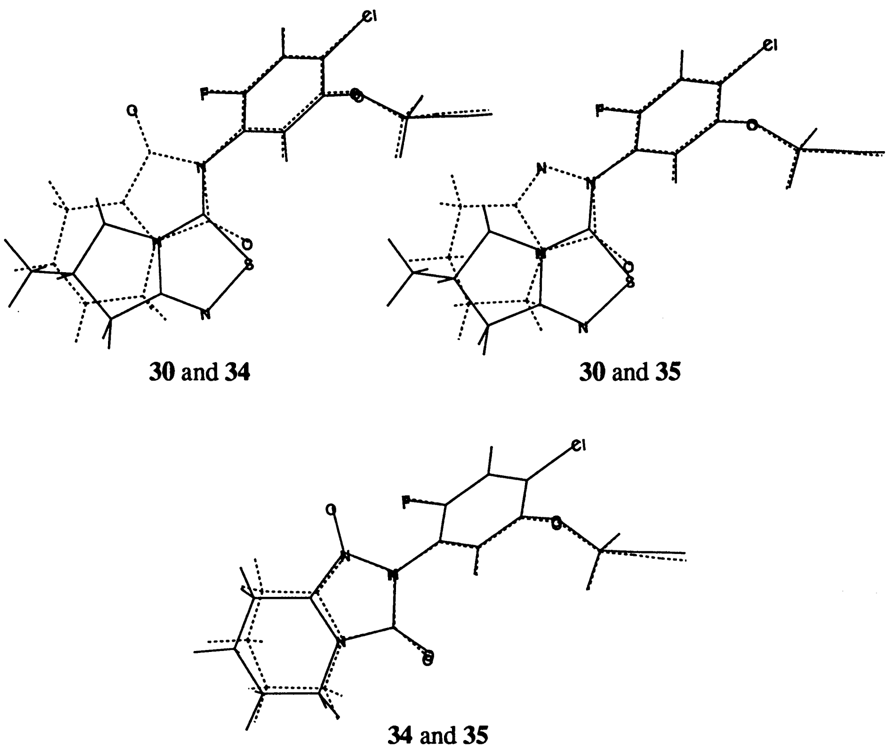
[b] Values from the superimposition by the least square fitting.

[c] Values from the optimized superimposition by the SIMPLEX fitting.

The R_{AB} values were in the range of 0.822 – 0.926 as shown in Table V, indicating that these molecules are in fact similar to each other with respect to the electrostatic properties. The SIMPLEX optimization procedure (*22*) was also employed to find the best superimposition of molecules to maximize the similarity index. The obtained superimposed structures and the values of optimized R_{AB} are shown in Figure 9 and Table V, respectively. Although the SIMPLEX R_{AB} values are slightly higher than those from the least square fitting, the degree of improvement is not large. The relative orientation of two compounds in each pair shown in Figure 9 was almost identical to the one by the least square fitting. The positions of the benzene ring and the alicyclic moiety in one compound were closely superimposed onto the corresponding positions of the other compound.

Steric Similarity: The molecular volume of compounds **30**, **34** and **35** were calculated and listed in Table VI. Their volume was in the range of *ca.* 264 – 275 Å^3. The volume occupied in common by the two of compounds superimposed as shown in Figure 9 was higher than 83% for each molecule. This indicates that these molecules are almost in an equivalent bulk, and their models superimposed to maximize the electrostatic similarity also satisfy their steric superimposability as well (*15*).

Structural Requirements for Light-dependent Herbicides: In the superimposed models of compound **30** onto compounds **34** and **35** in Figures 9a and b, the sulfur atom in the thiadiazoline ring of compound **30** is at the position of carbonyl oxygen in compounds **34** and **35**, suggesting that the sulfur may play a role similar to that of the carbonyl oxygen in compounds **34** and **35** in the herbicidal activity. The other carbonyl group in compound **34** and the nitrogen at the 1-position of compound **35** are not superimposed onto any atom in compound **30**, and the 2-nitrogen of compound **30** is also not on any atom in compounds **34** and **35**. One of the carbonyl groups in tetrahydrophthalimides (**IV**), the 2-nitrogen in 1,2,4-thiadiazolines (**I**) and the 1-nitrogen in 1,2,4-triazolin-3-ones (**V**) may not be essential for the activity. Their role may be to fix the molecular structure by fusing the alicyclic moiety and the hetero cyclic ring system. From the models shown in Figure 9, where the alicyclic moieties are closely superimposed on each other, a position-specificity of the hydrophobic alicyclic moiety in the molecule is expected to be important for the activity. The importance of the position-specific hydrophobicity is also understood



30 and 34

30 and 35

34 and 35

Figure 9. Superimposed models between two among three compounds **30**, **34** and **35** by the SIMPLEX optimization procedure (Reproduced with permission from ref. 15. Copyright 1994, Pestic. Sci. Society of Japan).

Table VI. Molecular Volume of Compounds 30, 34 and 35, and Their Common Volume in the Superimposed Models by the SIMPLEX Optimization[a]

Compd.	Molecular	Common volume (Å^3)[b]		
No.	Volume (Å^3)	30	34	35
30	273.8		236.1 (86.2)[c]	228.3 (83.4)[c]
34	274.9	236.1 (85.9)[c]		239.2 (87.0)[c]
35	263.5	228.3 (86.6)[c]	239.2 (90.7)[c]	

[a] Reproduced with permission from ref. 15. Copyright 1994, Pestic. Sci. Society of Japan.

[b] Calculated from the optimized superimposition shown in Figure 9.

[c] Values in parentheses are percentage of the common volume for each molecule.

because the presence of the methyl group on the alicyclic ring of compound **30** not only enhanced the herbicidal activity, but also improved significantly the superimposability with the alicyclic moieties of compounds **34** and **35**.

We propose the structural requirements for the activity as drawn in Figure 10. In this schematic representation, X is a hetero atom such as oxygen and sulfur, and Y is an sp^2 carbon or a planar nitrogen. At the position of Y, the sp^2 carbon of compound **34** and the planar nitrogens in compounds **30** and **35** are closely superimposed on each other as shown in Figure 9. The planar configuration around Y is considered to be important to fuse the alicyclic moiety with the heterocyclic ring coplanarly. The broken lines in Figure 10 represent two ways of the ring fusion to make the entire fused-ring moiety rigid. The "bioisosteric" structural requirements for light-dependent herbicides having dissimilar backbone structures such as **I**, **IV** and **V** could be understood through the hypothetical model shown in Figure 10 (*15*).

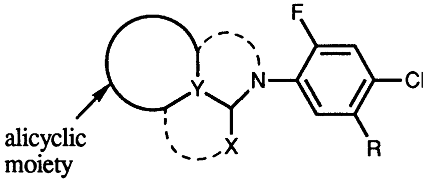


Figure 10. Schematic representation of the hypothetical structural requirements for the light-dependent herbicides (Reproduced with permission from ref. 15. Copyright 1994, Pestic. Sci. Society of Japan).

Conclusion

Two examples of the structure-activity study of the pesticidal 1,2,4-thiadiazolines were described above. In the case of the fungicidal Δ^3-1,2,4-thiadiazolines, the structure-activity relationship in a series of the congeneric compounds was quantitatively analyzed by the traditional QSAR procedure. The electron-density parameter derived from a quantum-chemical method was successful to clarify the importance of the chemical reactivity in governing the antifungal activity of the SH-inhibitors. In the second example for the herbicidal Δ^2-1,2,4-thiadiazolines, the molecular modeling techniques such as the similarity index and the conformation analysis were applied to extract the structural requirements of the light-dependent herbicides on the basis of the "dissimilar" structures including the non-congeneric cyclic imide and amide compounds. These studies were performed retrospectively after the synthetic project was over. Although the 1,2,4-thiadiazoline derivatives have not been developed as commercial pesticides, we believe that the methodologies and experiences during the process of the analyses will be extended in future studies to design new agrochemicals.

The acute toxicity of the fungicidal and herbicidal 1,2,4-thiadiazolines (**1** and **30**) was tested in our research center, and the LD_{50} values of these compounds against male mice were higher than 300 mg/Kg.

Acknowledgments

The authors are grateful to Professor Toshio Fujita for his invaluable suggestions and encouragement. Special thanks are due to Professor Corwin Hansch for providing us with the opportunity to publish our studies. Thanks are also given to many colleagues in Nippon Soda Company for their support to this work. The important aspects on the mode of action of herbicides were kindly provided by Professors Peter Böger of Universität Konstanz and Ko Wakabayashi of Tamagawa University.

References

1. Hagiwara, K.; Saitoh, K.; Iihama, T.; Kawana, T.; Hosaka, H. *J. Pesticide Sci.* **1993**, *18*, 309.
2. Hagiwara, K.; Hashimoto, S.; Shimoda, S. *J. Pesticide Sci.* **1992**, *17*, 251.
3. Hagiwara, K.; Hara, T.; Nomura, O.; Takahashi, H.; Hatano, R. *J. Pesticide Sci.* **1994**, *19*, 267.
4. Lukens, R. J.; Sisler, H. D. *Phytopathology* **1958**, *48*, 235.
5. Vincent, G.; Sisler, H. D. *Physiol. Plant.* **1968**, *21*, 1249.
6. Stewart, J. J. P. *J. Comput. Chem.*, **1989**, *10*, 209.
7. Stewart, J. J. P. *J. Comput. Chem.*, **1989**, *10*, 209.
8. Nakayama, A.; Hagiwara, K.; Hashimoto, S.; Shimoda, S. *Quant. Struct.-Act. Relat.* **1993**, *12*, 251.
9. Richards, W. G. *Quantum Pharmacology,* 2nd Ed.; Butterworths: Sevenoaks, U.K., 1983; pp 153-155.
10. Fukui, K.; Nagata, C.; Yonezawa, T. *J. Am. Chem. Soc.* **1958**, *60*, 2267.
11. Fujita, T. *Prog. Phys. Org. Chem.*, **1983**, *14*, 75.
12. Siegel, M. R.; Sisler, H. D. *Phytopathology*, **1968**, *58*, 1123.
13. Sikka, H. C.; Saxena, j.; Zweig, G. *Plant Physiol.* **1973**, *51*, 363.
14. Siegel, M. R.; Sisler, H. D. *Phytopathology*, **1968**, *58*, 1129.
15. Hagiwara, K.; Nakayama, A. *J. Pesticide Sci.* **1994**, *19*, 111.
16. Wakabayashi, K.; Böger, P. In *Pesticides-Environment: Molecular Biological Approaches;* Mitsui, T.; Matsumura, F.; Yamaguchi, I., Ed.; 1st Int. Symp. Pestic. Science; Pestic. Sci. Society of Japan: Tokyo, 1993, pp 239-253.
17. Watanabe, H.; Ohori, Y.; Sandmann, G.; Wakabayashi, K.; Böger, P. *Pestic. Biochem. Physiol.* **1992**, *42*, 99.
18. Carbo, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185.
19. Richards, W. G.; Hodgkin, E. E. *Chem. Br.* **1988**, 1141.
20. Hodgkin, E. E.; Richards, W. G. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 105.
21. Nakayama, A. *Quant. Struct.-Act. Relat.* **1992**, *11*, 478.
22. Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308.

RECEIVED July 18, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Chapter 17

Applications of a New Hydrophobicity Parameter of Amino Acid Side Chains to Quantitative Structure–Activity Analyses of Oligopeptides

Miki Akamatsu, Tamio Ueno, and Toshio Fujita[1]

Department of Agricultural Chemistry, Kyoto University, Kyoto 606–01, Japan

A new hydrophobicity parameter, π_α for the side chain of amino acid residues was defined by quantitatively analyzing the composition of experimentally measured log *P* value of oligopeptides and *N*-acetyl oligopeptide amides. It is comprized not only of the intrinsic π value of side chain substituents but also of other substituent factors to promote the aqueous/hydrophobic phase transfer of peptides. However, factors attributable to the conformational effects induced by intramolecular hydrogen-bonding such as β-turn and α-helix are not included in π_α. Structure-activity relationships for the platelet aggregation inhibition of the Arg-Gly-Asp-X (X : hydrophobic amino acid residue) series and for the opioid effects of two series of the gluten exorphin analogs, Tyr-Pro-X-Ser-Leu and Tyr-Pro-Ile-Gly-X (X : amino acid residue), were analyzed quantitatively using the π_α parameter and with others when necessary. The π_α parameter as the effective hydrophobicity index was shown to work nicely. The behaviors of outliers were reasonably explained by considering variations in the conformational equilibria between extended and β-turned forms in the hydrophobic environment.

Recently interests in peptidic and peptidomimetic bioactive compounds have been growing enormously (*1, 2*). To understand their distribution patterns in biological systems as well as their behaviors on the receptor level, one should know not only the hydrophobicity of the entire molecule but also that of the component amino acid side chains and peptide segments. Thus, we have been measuring the partition coefficient *P* in the 1-octanol/aqueous buffer system (pH 7) of a number of oligopeptides (*3, 4*) and *N*-acetyl oligopeptide amides (*5, 6*), the log *P* value being an index of the molecular hydrophobicity. We have analyzed the composition of log *P* values in terms of free-energy-related substituent and substructural parameters and defined an effective hydrophobicity constant, π_α, of the side chain substituents for natural amino acid residues under conditions in which the residues are contained in oligopeptides (*5*).

[1] Current address: EMIL Project, Fujitsu Kansai Systems Laboratory, 2–2–6 Shiromi, Chuo-ku, Osaka 540, Japan

0097–6156/95/0606–0229$12.00/0

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

The effective hydrophobicity constant, π_α, is a composite parameter comprized (in terms of a linear combination) of the following components: 1) the "intrinsic" substituent hydrophobicity, π, which one can estimate from the log P values of aliphatic compounds with and without the substituent in question under conditions in which any stereo-electronic effects of other nearby fragments can be ignored (1.02 π as one of the additive terms) (*7*, *8*), 2) a steric effect of the substituent which inhibits the relative solvation by 1-octanol of the nearby CONH groups in the peptide backbone. This effect is expressible by a Taft-type steric parameter, $E'_s{}^c$, where $E'_s{}^c$ is the "corrected" Dubois E'_s (*9*) representing not only the efficient bulk of substituents but also the effect of α-branching (0.23 $E'_s{}^c$ as one of the additive terms) (*10*), 3) a bidirectional electron withdrawing effect between the backbone CONH groups and the side-chain polar group so that the closer their distance, the higher becomes the increment in the log P value (as a specific value summed up with the component 4 assigned for each side chain), and 4) an intramolecular solvent-bridged hydrogen-bonding formation between the side-chain polar group and the backbone CONH to promote the partitioning in the hydrophobic phase (as a specific value summed up with the component 3 assigned for each side chain). The π_α parameter, in fact, represents the summation of various effects of each substituent on the partitioning of peptides into the 1-octanol from the aqueous phase other than effects attributable to the conformational factors brought about by intramolecular hydrogen-bonding such as β-turns and α-helices for higher peptides (*3*). Components 3 and 4 apply only to side chains carrying the polar groups.

For di- and tripeptides, the molecular hydrophobicity is represented by the summation of the side chain π_α values of constituent residues and components attributable to the backbone (*3*, *5*). For tetra- and higher peptides, additional components due to β-turns and α-helices are involved in the log P value (*4*). Thus, one of the applications of the π_α parameter is to obtain information about conformational situations such as β-turns and α-helices of higher peptides by comparing the measured log P value with that calculated in a way similar to those of di- and tripeptides. Another application of the π_α parameter is to use it to simulate the substituent hydrophobicity in the analyses of structure-activity relationship on the receptor level. Additional physicochemical effects including local conformational factors can be examined by adding other physicochemical parameter terms depending upon the situation.

In this chapter, we are going to present our attempts to quantitatively analyze the structure-activity relationships of oligopeptide series on the receptor level using the π_α constant and other free-energy related substituent parameters.

Applications to Structure-Activity Analyses

Platelet Aggregation Inhibition of Arg-Gly-Asp (RGD) Peptides. Platelet aggregation is the most important step in thrombus formation. It occurs through the binding of fibrinogen to its receptor, the platelet membrane glycoprotein GPIIb/IIIa, which recognizes fibrinogen by the sequence of RGD (*11*). Therefore, peptides containing the RGD sequence perhaps inhibit the binding of fibrinogen to GPIIb/IIIa and could be lead compounds of fibrinogen receptor antagonists. The introduction of a hydrophobic amino acid to the C-terminal of the Asp residue has been reported to enhance the inhibitory activity (*12*, *13*).

We synthesized a series of RGDX peptides (X = hydrophobic amino acids) and analyzed the relationship between the inhibitory activity and the hydrophobicity of X using the π_α value quantitatively (*14*). The activity of the test compounds was examined in the human platelet rich plasma (PRP) in terms of the inhibition of the ADP-mediated platelet aggregation. The 50 % inhibitory concentration (I_{50}, M) was measured for each compound. As the index for the inhibitory activity of RGDX, we used ΔpI_{50} which is the logarithm of the reciprocal of the I_{50} value relative to that of

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

the internal standard peptide Arg-Gly-Asp-Ser. The internal standard was included in each run of experimental sets, because the I_{50} value itself was not very constant depending upon individual blood samples. Table I shows the inhibitory activity ΔpI_{50}, and the π_α value and the structure of the side chain of the C-terminal residue X.

In Figure 1, the relationship between the ΔpI_{50} value and the π_α value of the residue X is depicted. The peptides seem to be roughly classified into two groups with two outliers. In the first peptide group, aromatic amino acids, Trp, Phe, Tyr and phenylglycine (Phg), and aliphatic amino acids, Val and Ser are attached as X. Peptides in the second group contain amino acids with other aliphatic side chains. The two outliers are peptides containing cyclohexylalanine (Cha) and *p*-chlorophenylalanine [F(*p*-Cl)]. In each group, the greater the hydrophobicity of the residue X, the higher the potency. Peptides having aliphatic Val and Ser side chains were included in the first group. The reason could be that alkyl groups in these side chains do not occupy the δ-position. The δ-methyl and methylene groups contained in the aliphatic side chains in the second group would be unfavorable for the binding of ligands to the receptor. The two outliers were less active than expected from their hydrophobicity. In the F(*p*-Cl), the eletron-withdrawing effect of Cl attached directly to the benzene ring could be deleterious to the activity. The Cha side chain, being aliphatic, has two endocyclic δ-methylene groups and its hydrophobicity may be too high for the binding interaction to occur properly. Using an indicator variable I_{alk} for Leu, Ile, Nva, and Nle, equation 1 was formulated.

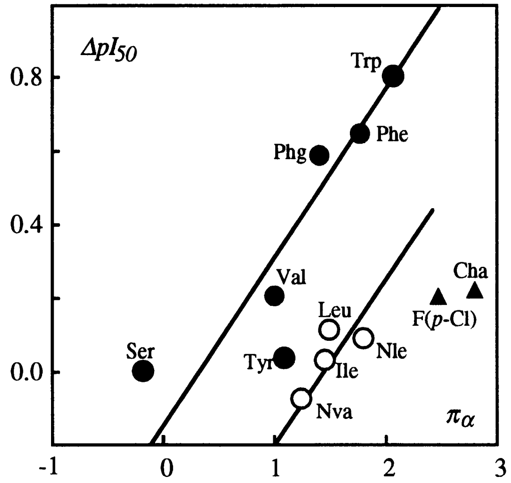


Figure 1. Relationship between Inhibitory Activity of RGDX and Hydrophobicity of Residue X

$$\Delta pI_{50} = 0.377(\pm 0.198)\, \pi_\alpha - 0.474(\pm 0.236)\, I_{alk} - 0.054\ (\pm 0.275) \qquad (1)$$

$$n = 10 \quad s = 0.149 \quad r = 0.911 \quad F_{2,7} = 17.1$$

In equation 1 and the following correlation equations, n is the number of compounds, s is the standard deviation, r is the correlation coefficient, F is the ratio of regression and residual variances, and the figures in parentheses are the 95 % confidence intervals.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Table I. Platelet Aggregation Inhibitory Activity of Arg-Gly-Asp-X (RGDX) and the π_α Value of the Residue X

No.	Peptides		π_α	I_{alk}	ΔpI_{50}		
	X	Side Chain of X			Obsd.[a]	Calcd.[b]	Δ
	RGDX						
1	Ser	$-CH_2OH$	-0.19	0	0.00	-0.08	0.08
2	Val	$-CH(CH_3)CH_3$	1.00	0	0.27	0.35	-0.08
3	Leu	$-CH_2CH(CH_3)CH_3$	1.52	1	0.18	0.01	0.17
4	Ile	$-CH(CH_3)CH_2CH_3$	1.43	1	0.00	-0.02	0.02
5	Phe	$-CH_2Ph$	1.78	0	0.65	0.63	0.02
6	Tyr	$-CH_2Ph(4\text{-}OH)$	1.10	0	0.06	0.38	-0.32
7	Trp	Indole-3-methyl	2.02	0	0.80	0.71	0.09
8	Nva[c]	$-CH_2CH_2CH_3$	1.24[d]	1	-0.15	-0.09	-0.06
9	Nle[c]	$-CH_2CH_2CH_2CH_3$	1.79[d]	1	0.11	0.11	0.00
10	Phg[c]	-Ph	1.42[e]	0	0.58	0.50	0.08
11	Cha[c]	$-CH_2$-*cyc*-C_6H_{11}	2.81[f]	0	0.22[k]	0.99	-0.77
12	F(*p*-Cl)[c]	$-CH_2Ph(4\text{-}Cl)$	2.50[g]	0	0.20[k]	0.88	-0.68
	cyclic-Peptides						
13	Leu	$-CH_2CH(CH_3)CH_3$	1.52	1	1.74	1.87	-0.13
14	Val	$-CH(CH_3)CH_3$	1.00	0	2.38	2.21	0.17
15	Phe	$-CH_2Ph$	1.78	0	2.43	2.49	-0.06
16	Tyr	$-CH_2Ph(4\text{-}OH)$	1.10	0	2.23	2.24	-0.01
17	Y(Me)[c]	$-CH_2Ph(4\text{-}OMe)$	1.76[h]	0	2.47	2.48	-0.01
18	Y(*n*Bu)[c]	$-CH_2Ph(4\text{-}O\text{-}n\text{-}Bu)$	3.41[i]	0	2.68[k]	3.07	-0.39
19	Phg[c]	-Ph	1.42[e]	0	2.39	2.36	0.03
20	F(*p*-I)[c]	$-CH_2Ph(4\text{-}I)$	2.92[j]	0	2.37[k]	2.89	-0.52
21	F(*p*-Cl)[c]	$-CH_2Ph(4\text{-}Cl)$	2.50[g]	0	2.36[k]	2.74	-0.38

[a] The pI_{50} value relative to that of Arg-Gly-Asp-Ser (1), being 3.78(±0.10), (n = 9).

[b] Calculated by equation 2. [c] Nva : norvaline; Nle : norleucine; Phg : phenylglycine; Cha : cyclohexylalanine; F(*p*-Cl) : *p*-chlorophenylalanine; Y(Me) : *O*-methyltyrosine; Y(*n*Bu) : *O*-*n*-butyltyrosine; F(*p*-I) : *p*-iodophenylalanine. [d] π_α(Nva) = 1.02 π(*n*-Pr) + F_{gBr} + 0.23 $E'_s{}^c$(*n*-Pr) = 1.02 x 1.62 +(-0.22) + 0.23 x (-0.82) =1.24. π_α(Nle) = 1.02 π(*n*-Bu) + F_{gBr} + 0.23 $E'_s{}^c$(*n*-Bu) = 1.02 x 2.16 +(-0.22) + 0.23 x (-0.82) =1.79. For the coefficient of the π term, 1.02, see ref. *6*. F_{gBr} is a factor for the group branching at the α-position of amino acid residues. See Hansch, C.; Leo, A.J. *Substitution Constants for Correlation Analysis in Chemistry and Biology*; John Wiley: New York, NY, 1979; pp 18-43. [e] Newly estimated from the log *P* value of *N*-acetyl di- and tripeptide amides containing a Phg residue. [f] From ref. 27. [g] π_α[F(*p*-Cl)] = π_α(Phe) + 1.02 π(Cl) = 1.78 + 1.02 x 0.71 = 2.50. The $E'_s{}^c$[F(*p*-Cl)] value was taken as that of Phe.

[h] π_α[Y(Me)] = π_α(Phe) + 1.02 π(OMe) = 1.78 + 1.02 x (-0.02) = 1.76. The $E'_s{}^c$[Y(Me)] value was taken as that of Phe. [i] π_α[Y(*n*Bu)] = π_α[Y(Me)] + 1.02 x [3 x π(Me)] = 1.76 + 1.02 x 3 x 0.54 = 3.41. The $E'_s{}^c$[Y(O*n*Bu)] value was taken as that of Phe. [j] π_α[F(*p*-I)] = π_α(Phe) + 1.02 π(I) = 1.78 + 1.02 x 1.12 = 2.92. The $E'_s{}^c$[Y(*n*Bu)] value was taken as that of Phe. [k] Excluded from the analysis of equation 2.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Recently, a series of cyclic peptides [Arg(Pmc)-Gly-His-Arg-Gly-Asp-X-Arg-
S——————————
Cys-Arg] has been reported to be very potent fibrinogen antagonists (*12*). These
—S
peptides also contain the RGDX sequence, in which X is the hydrophobic residue such as Leu, Val, Phe, Tyr, Y(Me), Y(*n*Bu), Phg, F(*p*-Cl), and F(*p*-I) [Y(R) and F(R) are *O*-R-tyrosine and R-phenylalanine, respectively]. Introduction of aromatic hydrophobic amino acids and the shorter aliphatic residue Val at the X position are shown to be more effective in potentiating the activity. The structure-activity pattern is similar to that for the linear RGDX peptides. Thus, we analyzed the ΔpI_{50} values for the combined set with our peptides to formulate equation 2.

$$\Delta pI_{50} = 0.356(\pm 0.151)\,\pi_\alpha - 0.522(\pm 0.167)\,I_{alk} + 1.859(\pm 0.158)\,I_{cyc} - 0.009(\pm 0.215) \qquad (2)$$

$$n = 16 \quad s = 0.134 \quad r = 0.994 \quad F_{3,12} = 305$$

In equation 2, I_{cyc} is another indicator variable for cyclic peptides. The cyclic peptides in which X = F(*p*-Cl), F(*p*-I), and Y(*n*Bu) were outliers and not included in equation 2. Their potency is lower than expected from the hydrophobicity of the side chain. For outlying behavior of the F(*p*-Cl) and F(*p*-I) analogs, the same reasoning as that for the F(*p*-Cl) in the linear peptides could be anticipated. In the Y(*n*Bu) analog, although its potency is high, the substituent would be so lengthy or hydrophobic that the potency seems to be leveled off beyond that of the lower homolog, perhaps, Y(Et). Except for the addition of the I_{cyc} term, the coefficients of the corresponding terms in equation 2 are very similar to those in equation 1, suggesting that the RGDX sequence of both linear and cyclic peptides binds to the receptor in a similar fashion to each other. The effect of the aromatic and the shorter aliphatic side chains relative to that of the longer aliphatic ones is almost identical between linear and cyclic peptide series. The correlation result is shown in Table I.

We have shown empirically that the conformation of linear tetra- and pentapeptides in the aqueous phase is mostly extended and "random", but the β-turn formation is significant in the more hydrophobic environment such as in the 1-octanol phase (*5*). The ease of the β-turn formation of four consecutive amino acid residues is determined by the propensity of each amino acid at each position to participate in this conformational modification. In the present case, the propensity of the fourth amino acids could be the determinant. The β-turn propensity parameter defined by Chou and Fasman (*15*) for the X residue at the fourth position of the β-turn (not shown) was found not significant when included in equation 1. The RGD unit has been reported to take nearly extended conformation in a snake venom, kistrin (*16*), as well as in cyclic peptides such as SK&F 106760 (*17*). The slope of the π_α term, about 0.36, suggests that approximately one third of the surface of the side chain substituents is desolvated upon the receptor binding. The "hydrophobic interaction" of the X-residues does not seem to be an engulfed type.

Initially, we were skeptical in classifying the linear RGDX peptides into two groups so that two compounds having Ser and Val residues are separated from others having alkyl side chains. Although the physicochemical meaning of the I_{alk} parameter is still not clear, the relevancy of the classification seemed to be reinforced by the structure-activity pattern of cyclic peptides.

Opioid Activity of Gluten Exorphin Analogs. The presence of opioid activity has been recognized in the peptic digest of wheat gluten (*18, 19*). Fukudome and Yoshikawa isolated five novel opioid peptides from the enzymatic digests of gluten and named them gluten exorphin A4, A5, B4, B5 and C (*20*). Gluten exorphin C (Tyr-Pro-Ile-Ser-Leu) is quite different from any of the reported exogenous opioid peptides such as β-casomorphin 7 (Tyr-Pro-Phe-Pro-Gly-Pro-Ile), hemorphin (Tyr-Pro-Trp-Thr-Gln), cytochrophin (Tyr-Pro-Phe-Thr-Ile) in the fact that it has no aromatic amino acid but the hydrophobic Ile at the position 3 following the Tyr-Pro sequence. Thus, the effect of the amino acid replacement at the position 3 of exorphin C on the opioid activities was investigated using synthetic analogs shown in Table II (Tyr-Pro-X-Ser-Leu, X = L-amino acids : **1-7**) (*21*). The opioid activities were measured in terms of the concentration of peptides, in which strips of guinea pig ileum longitudinal muscle (GPI) and mouse vas deferens (MVD) are immersed under certain conditions, to reduce the electrically induced smooth muscle contractions to 50% of the control (I_{50}, M).

A fragment peptide of wheat gliadin, Tyr-Pro-Leu-Gly-Gln (**8**) which contains the Tyr-Pro-X sequence was also isolated and named gluten exorphin D. It showed a higher activity for the GPI receptor than gluten exorphin C (*21*). Based on this observation, Fukudome and Yoshikawa synthesized a series of hybrid analogs of gluten exorphins C and D which have the sequence Tyr-Pro-Ile-Gly-X (X = L-amino acids : **9-17**), and examined the effect of the amino acid replacements at the position 5. In this series of peptides, the Ile^3 of exorphin C and the Gly^4 of exorphin D were retained, because Tyr-Pro-Ile-Gly-Gln (**9**) and Tyr-Pro-Ile-Gly-Leu (**10**) showed the higher potency than exorphin D, Tyr-Pro-Leu-Gly-Gln (**8**), and exorphin C, Tyr-Pro-Ile-Ser-Leu (**1**), respectively, in the GPI and MVD test systems. Table II shows the opioid activities and side chain hydrophobic and steric parameters of gluten exorphin analogs.

First, we examined the relationship between two activities for the GPI and MVD receptors of exorphin analogs. The existence of at least three types of opioid receptors which were named μ, δ, and κ has been shown (*22*). The GPI and MVD preparations are known to contain μ- and δ-receptors, respectively. For the Tyr-Pro-Ile-Gly-X (**9-16**) and Tyr-Pro-X-Ser-Leu (**1-7**) series, equations 3 and 4 were formulated, respectively.

$$pI_{50}(\text{GPI}) = 0.611(\pm0.301)\, pI_{50}(\text{MVD}) + 2.815(\pm1.426) \tag{3}$$
$$n = 8 \quad s = 0.252 \quad r = 0.897 \quad F_{1,6} = 24.8$$

$$pI_{50}(\text{GPI}) = 0.488(\pm0.474)\, pI_{50}(\text{MVD}) + 0.681(\pm0.707)\, \pi_3 + 1.131(\pm2.215) \tag{4}$$
$$n = 7 \quad s = 0.218 \quad r = 0.902 \quad F_{2,4} = 8.68$$

In equation 4, π_3 is the π_α value of the third residue in the series of Tyr-Pro-X-Ser-Leu, the term being justified by the t-test at the 94.4% level of significance. The slope of the pI_{50}(MVD) terms in these equations is similar to each other and it indicates that the susceptibility of the GPI receptors is about a half that of the MVD receptors to the corresponding structural variations in both series of peptides, if the extra contribution of the hydrophobicity in the Tyr-Pro-X-Ser-Leu series is separated by the π_3 term. Thus, equations 3 and 4 are combinable yielding equation 5 using an indicator variable I_3 for the Tyr-Pro-X-Ser-Leu series and incorporating the π_3 value of the Ile side chain for the Tyr-Pro-Ile-Gly-X series.

$$pI_{50}(\text{GPI}) = 0.577(\pm0.211)\, pI_{50}(\text{MVD}) + 0.666(\pm0.591)\, \pi_3 - 1.254(\pm0.283)\, I_3 + 2.026(\pm1.273) \tag{5}$$
$$n = 15 \quad s = 0.231 \quad r = 0.971 \quad F_{3,11} = 60.0$$

In equation 5, the correlation quality was improved and the level of significance of the π_3 term was elevated. The π_3 term indicates that the hydrophobic interaction of the third residue with the GPI (μ) receptor is more important than that with the MVD (δ) receptor in the Tyr-Pro-X-Ser-Leu series of peptides. As to whether this extra hydrophobic interaction is significant for the Ile residue in the Tyr-Pro-Ile-Gly-X series, however, no definite conclusion can be drawn because its contribution

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Table II. Opioid Activities and Physicochemical Side Chain Parameters of Gluten Exorphin Analogs

No.	Peptides[a]	π_3	π_5	$E'_s{}^c(3)$[b]	pI_{50}							
					GPI					MVD		
					Obsd.[c]	Calcd.				Obsd.[c]	Calcd.	
						Eq. 5	Δ	Eqs. 6, 7[d]	Δ		Eqs. 8, 9[e]	Δ
1	YPISL	1.43	1.52	-1.81	4.40[f]	4.53	-0.13	3.16	1.24	4.87[g]	2.98	1.89
2	YPLSL	1.52	1.52	-1.44	3.70	3.92	-0.22	3.69	0.01	3.70	3.57	0.13
3	YPVSL	1.00	1.52	-1.29	3.70	3.57	0.13	3.72	-0.02	3.70	3.80	-0.10
4	YPNleSL[h]	1.79	1.52	-0.82	4.52	4.74	-0.22	4.62	-0.10	4.82	4.55	0.27
5	YPNvaSL[h]	1.24	1.52	-0.82	4.48	4.40	0.08	4.44	0.04	4.85	4.55	0.30
6	YPFSL	1.78	1.52	-0.90	4.52	4.35	0.17	4.51	0.01	4.15	4.42	-0.27
7	YPWSL	2.02	1.52	-0.86	4.70	4.51	0.19	4.65	0.06	4.15	4.49	-0.34
8	YPLGQ	1.52	-0.52	-1.44	5.15[i]	5.58	-0.43	-	-	4.41	-	-
9	YPIGQ	1.43	-0.52	-1.81	5.52[j]	5.73	-0.21	5.11	0.41	4.77[k]	4.05	0.72
10	YPIGL	1.43	1.52	-1.81	6.26	6.05	0.21	6.15	0.11	5.33	5.36	-0.03
11	YPIGI	1.43	1.43	-1.81	6.06	6.02	0.04	6.08	-0.02	5.27	5.28	-0.01
12	YPIGF	1.43	1.78	-1.81	6.44	6.31	0.13	6.32	0.12	5.77	5.59	0.18
13	YPIGV	1.43	1.00	-1.81	5.49	5.73	-0.24	5.79	-0.30	4.78	4.91	-0.13
14	YPIGA	1.43	0.28	-1.81	5.38	5.44	-0.06	5.29	0.09	4.26	4.28	-0.02
15	YPIGK	1.43	-1.87	-1.81	5.46	5.08	0.38	5.25	0.21	3.64	3.90	-0.26
16	YPIGE	1.43	-2.39	-1.81	4.85	5.11	-0.26	4.89	-0.04	3.70	3.44	0.26
17	YPIGR	1.43	-1.31	-1.81	5.46	-	-	5.63	-0.17	-	-	-

[a] Amino acids are written by the one-letter representation, unless otherwise noted. [b] The value for the side chain of the third residue. From ref. 10. [c] The I_{50} values were taken from ref. 21. [d] Calculated by equation 7 for peptides **2-7** and by equation 6 for peptides **10-17**. [e] Calculated by equation 9 for peptides **2-7** and by equation 8 for peptides **10-17**. [f] Excluded from the analysis of equation 7. [g] Excluded from the analysis of equation 9. [h] Nle: norleucine; Nva: norvaline. [i] Excluded from the analysis of equation 5. [j] Excluded from the analysis of equation 6. [k] Excluded from the analysis of equation 8.

represented as π_3(Ile) in equation 5 is counterbalanced by the I_3 and constant terms. In this respect, it should be noted that the pI_{50} value of exorphin D, Tyr-Pro-Leu-Gly-Gln (**8**), which belongs to neither of the two series of peptides analyzed here, is not very far apart from that calculated (5.58) by equation 5 with consideration of π_3 of the Leu side chain and $I_3 = 0$. The contribution of the π_3 hydrophobicity may not totally be ignored here.

Next, we analyzed the activities for two types of receptor preparations with free energy-related physicochemical parameters for the side chain substituents. The activity for the GPI system of the series Tyr-Pro-Ile-Gly-X (**10-17**) was linearly related with the hydrophobicity π_α of the fifth residue, π_5, equation 6 being formulated.

$$pI_{50}(\mathrm{GPI}) = 0.689(\pm0.367)\ \pi_5 + 1.437(\pm1.184)\ I_{ion} + 5.098(\pm0.497) \qquad (6)$$
$$n = 8 \quad s = 0.200 \quad r = 0.948 \quad F_{2,5} = 22.1$$

In equation 6, Tyr-Pro-Ile-Gly-Gln (**9**) is not included because its activity is much higher than expected from the π_5(Gln) value. For the ionizable side chains, Lys, Asp, and Arg (in peptides **15-17**), the π_α values for their ion-paired form under physiological conditions were used as π_5 with a correction along with an indicator variable I_{ion} (*6*). To define the π_α value for ionizable side chains (*23*), the "apparent" hydrophobicity π' value is based on the apparent log P (log P') vlaue of the ion-paired

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

peptides at pH 7 (*6*). We have used $C_3H_7NH_3^+$ as the countercation for acidic aspartyl and glutamyl peptides, and $C_2H_5COO^-$ as the counteranion for basic arginyl and lysyl peptides. Acidic and basic side chains form the same type of ion-pairs, $R\text{-}COO^- \cdot H_3N^+\text{-}R'$ (*6*). Because this series of peptides exhibit the biological activity regardless of whether the X residue is basic, neutral or acidic, ionized functions in basic and acidic peptides seem to interact with neither anionic nor cationic "receptor sites" specifically. Therefore, considering that the alkyl chain in the propylammonium ion as the counterion of the acidic side chain is one methylene unit longer than that in the propionate anion as the counterion of the basic side chain, the π_α value of the ion-pair of the acidic side chain with the countercation was corrected in such a way that $\pi_\alpha(\text{Glu})_{\text{cor}} = \pi_\alpha(\text{Glu}) - \pi(CH_3) = -1.85 - 0.54 = -2.39$ (for π_α values, see ref. *6*), and used as the π_5 value of the Glu side chain to make the standard the same as the π_α for basic side chains.

For the series Tyr-Pro-X-Ser-Leu (**2**-**7**), not only the hydrophobic effect in terms of π_3 but also steric effect in terms of $E'_s{}^c$ of the third residue was significant, as shown in equation 7, in which, however, exorphin C (**1**) is not included.

$$pI_{50}(\text{GPI}) = 0.338(\pm 0.308)\,\pi_3 + 1.353(\pm 0.434)\,E'_s{}^c(3) + 5.126(\pm 0.801) \qquad (7)$$
$$n = 6 \quad s = 0.073 \quad r = 0.992 \quad F_{2,3} = 93.8$$

Equations 8 and 9 for the pI_{50}(MVD) value were obtained as the counterparts of equations 6 and 7, respectively. Again, the peptide **9** was omitted in equation 8, so was exorphin C (**1**) in equation 9.

$$pI_{50}(\text{MVD}) = 0.874(\pm 0.482)\,\pi_5 + 1.500(\pm 1.683)\,I_{ion} + 4.031(\pm 0.638) \qquad (8)$$
$$n = 7 \quad s = 0.214 \quad r = 0.978 \quad F_{2,4} = 43.8$$

$$pI_{50}(\text{MVD}) = 1.585(\pm 1.407)\,E'_s{}^c(3) + 5.848(\pm 1.479) \qquad (9)$$
$$n = 6 \quad s = 0.308 \quad r = 0.842 \quad F_{1,4} = 9.78$$

In equation 8, only two peptides with ionizable side chains, Lys (in **15**) and Asp (in **16**), were included in the analysis. Thus, the I_{ion} term was justified only at the 93.1% level. The coefficients of the π_5 and I_{ion} terms are, however, higher than the corresponding values in equation 6, conforming to equation 3. The slope of the $E'_s{}^c$ term in equation 9 is larger than that in equation 7. This along with the fact that the addition of the π_3 term into equation 9 is insignificant is not inconsistent with equations 4 and 5.

The peptide **1** (exorphin C) was accommodated well in equations 4 and 5. But, it was the outlier from equations 7 and 9. The pI_{50} values of this compound for the GPI and MVD prepartions were much higher than those calculated by eqations 7 and 9. Thus, its outlying behaviors were likely to arise from a common physicochemical origin. As described in the preceding section, pentapeptides dealt with here are able to take the β-turned conformation in the hydrophobic environment. The β-turn formation is mostly due to the hydrogen-bond formation between the CO-oxygen of the (i)th residue and the NH-hydrogen of the (i+3)th residue in four consecutive amino acid sequence in peptides. At the four consecutive positions, the propensity of the Pro residue is among the highest at the (i+1)th (second) position so that the exorphin C analogs, Tyr-Pro-X-Ser-Leu, were considered to exist as the β-turned structure almost exclusively within the Tyr-Pro-X-Ser sequence in the receptor-bound state which is regarded as being more hydrophobic than in the bulk aqueous phase. The overall β-turn propensity of Tyr-Pro-X-Ser sequence is dependent on the propensity of the (i+2)th (third) residue X. We have examined possible linear free-energy relationship between the propensity index of residues and physicochemical parameters of the side chain substituents and found that a "free-energy-related" index

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

of the propensity based on that of Chou and Fasman (not shown) for a certain amino acid to exist at the (i+2)th position in the β-turned sequence is linearly related with the $E'_s{}^c$ parameter of the side chain substituent of the amino acid (*4*). By definition, the bulkier as well as the more α-branched the substituents, the more negative the $E'_s{}^c$ value. That is, the bulkier as well as the more α-branched the side chain substituent at the (i+2)th position, the greater is the twisting of the CO(i+2) - NH(i+3) bond so that the direction of the NH group of the (i+3)th residue is more severely distorted and more difficult to form the intramolecular hydrogen-bonding with the CO group of the (i)th residue. Thus, the propensity for that residue to participate in the β-turn formation at the (i+2)th position will be lower. The positive $E'_s{}^c$ term in equations 7 and 9 strongly suggests that the lower the bulkiness of the side-chain substituent at the (i+2)th position, the higher is the β-turn propensity in the Tyr-Pro-X-Ser sequence. Except for exorphin C itself, the exorphin C analogs seem to interact with the GPI as well as the MVD receptors as the β-turned conformation of the Tyr-Pro-X-Ser sequence. For the GPI receptors, the hydrophobicity in terms of π_α of the side chain of the X residue works to promote the affinity as shown in equation 7.

In exorphin C, X is Ile, the Chou-Fasman's β-turn propensity of which is among the lowest as the (i+2)th residue (data not shown). Thus, compensating the very high propensity of the (i+1)th Pro residue, the Ile residue in exorphin C tends to maintain the conformation of the Tyr-Pro-Ile-Ser sequence not β-turned. For the alternative sequence, Pro-Ile-Ser-Leu, no significant propensity to take the β-turn is anticipated. Thus, the receptor-binding of exorphin C seems to occur with the extended and "random" conformation differing from its analogs. Although the three-dimensional structures of peptides in relation to their receptor-selective binding have not adequately been understood, it has been reported that both the β-turned monomer and the extended dimer of natural enkephalins can bind with both the μ- and δ-types of the opioid receptor (*24, 25*). There seems to exist the highest "bulk" for the side-chain substituent at the (i+2)th position beyond which the propensity for the peptides to take the β-turned conformation is insignificant. The above argument seems to rationalize the outlying behaviors of exorphin C, which is quite active, from equations 7 and 9.

In the peptide series of Tyr-Pro-Ile-Gly-X (**9**-**17**), peptide **9** in which X is Gln showed an outlyingly higher activity than that calculated by equations 6 and 8. The peptide **9** was well accommodated in equations 3 and 5, so that a common reasoning for its behaviors was anticipated. This series of peptides is also considered to exist as an equilibrium mixture between extended and β-turned conformers in the hydrophobic milieu. Because of the very high β-turn propensity of the Pro residue at the (i+1)th position and the Gly residue at the (i+3)th position as well, this series of peptides is also considered to take the β-turned conformation in spite of the oppositely working effect of the Ile residue at the (i+2)th position within the Tyr-Pro-Ile-Gly sequence. Exception may be the peptide, Tyr-Pro-Ile-Gly-Gln (**9**), because the β-turn propensity of the Gln as the (i+3)th residue in the Pro-Ile-Gly-Gln sequence is among the highest. Even though the participation of the β-turned form of the Tyr-Pro-Ile-Gly sequence still exists, that of the Pro-Ile-Gly-Gln cannot be ignored. The second conformation could show a higher potency than the first for the GPI and MVD receptors. The above argument may rationalize the outlying behavior of the peptide **9**.

In equations 6 and 8, the hydrophobicity of the fifth residue appears to indicate that the extent of the desolvation is greater than 0.5, suggesting the side chain substituent at this position is somewhat engulfed inside the receptor cavity. It should be noted that the size of the I_{ion} term in equations 6 and 8 has no concrete physicochemical meaning, because the π_5 value for the ionizable side chains has a standard different from the π_5 value for the neutral side chains. The use of the I_{ion} term is just to adjust this difference.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Discussion

Our new index, π_α, is distinctively different from most of the hydrophobicity indices of amino acid side chains published so far in that it is estimated from the experimentally measured "net" hydrophobicity of oligopeptides existing as such in solutions. It is an "effective" substituent parameter including not only the "intrinsic" hydrophobicity but also other factors contributing to the transfer of peptides from aqueous to hydrophobic phase under conditions in which the side chain substituents are included in peptidic molecules. It does not include, however, components which arise from conformational factors which accompany the intramolecular hydrogen-bonding observed in those higher than tripeptides. Thus, the participation of the conformational factors could be examined separately.

A set of π values for the side chain of amino acids has been proposed by Fauchere and Pliska (*26*). Their values are based upon the log *P* value of *N*-acetylamino acid amides measured at pH 7. Although they are not based on the log *P* value of peptides, the Fauchere-Pliska π values are similar to our π_α values in that they seem to not only count the intrinsic π value of substituents but also include components similar to those considered for our π_α value. In fact, equation 10 was formulated between their π and our π_α values for non-ionizable side chains with a high correlation (Akamatsu, M.; Fujita, T. In *QSAR : New Developments and Applications*; Fujita, T., Ed.; Elsevier Science Publishers: Amsterdam, in press).

$$\pi(\text{Fauchere-Pliska}) = 1.254(\pm 0.175)\,\pi_\alpha - 0.010(\pm 0.167) \qquad (10)$$
$$n = 13 \quad s = 0.198 \quad r = 0.979 \quad F_{1,11} = 248$$

Fauchere and Pliska did not pay special attention to ionizable side chains, and their π values are simply calculated from the apparent log *P* values measured at pH 7, as ajusted with 0.1 *N* HCl or NaOH. Therefore, for the structure-activity analysis of platelet-aggregation inhibition of RGDX peptides in which no side chain of the X residues was ionizable, the correlation results with their π value were almost equivalent with those of equations 1 and 2 (data not shown). For the structure-activity analysis of gluten exorphin analogs, however, the direct use of Fauchere-Pliska π value for the data set including peptides having ionizable X side chain without the I_{ion} term yielded corresponding correlations poorer than equation 6 and 8. The counterpart of equation 6 showed $s = 0.352$ and $r = 0.789$ and that of equation 8, $s = 0.221$ and $r = 0.970$. Although the quality of the latter is not much lower than that of equation 8, this should be regarded as being fortuitous because the standardization of the scale of π values for acidic and basic side chain substituents in the Fauchere-Pliska system is not made explicitly.

In conclusion, it should be noted that the two examples of the application of our π_α system to quantitative structure-activity analyses of oligopeptide series represented above have by no means been completed. They should be elaborated with additional compounds to draw definite conclusions. The number of compounds included in many of the correlation equations is fewer than one might like, but the equations seem physicochemically not unreasonable. Moreover, rationalizations of outliers as well as combinations of structure-activity patterns for "different" series of analogs using indicator variables were believed to be carried out characteristically for the peptidic compounds. We have shown that the molecular hydrophobicity of cyclic decapeptide gramicidin S analogs can be analyzable with the π_α parameter modified for a difference in the steric effect of side chain substituents described as the component 2 earlier in this article (*27*). We hope that the π_α system will be utilized more often so that its versatility in the field of the structure-activity analyses of peptidic and peptidomimetic bioactive compounds can be established.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch017

Acknowledgments

We thank Professor Masaaki Yoshikawa of the Department of Food Science and Technology for his helpful suggestions.

Literature Cited

1. Dutta, A.S. *Small Peptides: Chemistry, Biology and Clinical Studies*; Pharmacochemistry Library 19; Elsevier Science Publishers: Amsterdam, 1993.
2. Nachman, R.J.; Tilley, J.W.; Hayes, T.K.; Holman, G.M.; Beier, R.C. In *Natural and Engineered Pest Management Agents*; Hedin, P.A.; Menn, J.J.; Hollingworth, R.M., Eds.; ACS Symposium Series 551; American Chemical Society: Washington, D.C., 1994; pp 210-229.
3. Akamatsu, M.; Yoshida, Y.; Nakamura, H.; Asao, M.; Iwamura, H.; Fujita, T. *Quant. Struct.-Act. Relat.* **1989**, *8*, 195-203.
4. Akamatsu, M.; Fujita, T. *J. Pharm. Sci.* **1992**, *81*, 164-174.
5. Akamatsu, M.; Okutani, S.; Nakao, K., Hong, N.J.; Fujita, T. *Quant. Struct.-Act. Relat.* **1990**, *9*, 189-194.
6. Akamatsu, M.; Katayama, T.; Kishimoto, D.; Kurokawa, Y.; Shibata, H.; Ueno, T.; Fujita, T. *J. Pharm. Sci.* **1994**, *83*, 1026-1033.
7. Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175-5180.
8. Iwasa, J.; Fujita, T.; Hansch, C. *J. Med. Chem.* **1965**, *8*, 150-153.
9. MacPhee, J.A.; Panaye, A.; Dubois, J.-E. *Tetrahedron* **1978**, *34*, 3553-3562.
10. Takayama, C.; Akamatsu, M.; Fujita, T. *Quant. Struct.-Act. Relat.* **1985**, *4*, 149-160.
11. Ruoslahti, E.; Pierschbacher, M.D. *Science* **1987**, *238*, 491-497.
12. Charon, M.H.; Poggi, A.; Donati, M.B.; Marguerie, G. In *Peptides*, Rivier, J.E.; Marshall, G.R., Eds.; ESCOM Science Publishers B.V.: Leiden, **1990**, pp 82-83.
13. Cheng, S.; Craig, W.S.; Mullen, D.; Tschopp, J.F.; Dixon, D.; Pierschbacher, M.D. *J. Med. Chem.* **1994**, *37*, 1-8.
14. Miyashita, M.; Ueno, H.; Akamatsu, M.; Nishimura, K.; Ueno, T.; Fujita, T. In *Peptide Chemistry 1993*; Okada, Y., Ed.; Protein Research Foundation: Osaka, **1994**; pp 241-244.
15. Chou, P.Y.; Fasman, G.D. *J. Mol. Biol.* **1977**, *115*, 135-175.
16. Alder, M.; Wagner, G. *Biochemistry* **1992**, *31*, 1031-1039.
17. Peishoff, C.E.; Ali, F.E.; Bean, J.W.; Calvo, R.; D'Ambrosio, C.A.; Eggleston, D.S.; Hwang, S.M.; Kline, T.P.; Koster, P.F.; Nichols, A.; Powers, D.; Romoff, T.; Samanen, J.M.; Stadel, J.; Vasko, J.A.; Kopple, K.D. *J. Med. Chem.* **1992**, *35*, 3962-3969.
18. Zioudrou, C.; Stereaty, R.C.; Klee, W.A. *J. Biol. Chem.* **1979**, *254*, 2446-2449.
19. Huebner, F.R.; Lieberman, K.W.; Rubino, R.P.; Wall, J.S. *Peptides* **1984**, *5*, 1139-1147.
20. Fukudome, S.; Yoshikawa, M. *FEBS Lett.* **1992**, *296*, 107-111.
21. Fukudome, S.; Yoshikawa, M. In *Peptide Chemistry 1992*; Yanaihara, N., Ed.; ESCOM Science Publishers B. V.: Leiden, **1993**, pp 390-392.
22. Gilbert, P.E.; Martin, W.R. *J. Pharmacol. Exp. Ther.* **1976**, *198*, 66-82.
23. In the original publication (ref. 6), π_α for ionizable side chains was defined as π_α'.
24. Doi, M.; Tanaka, M.; Ishida, T.; Inoue, M.; Fujiwara, T.; Tomita, K.; Kimura, T.; Sakakibara, S.; Sheldrick, G.M. *J. Biochem.* **1987**, *101*, 485-490.
25. Doi, M.; Tanaka, M.; Ishida, T.; Inoue, M. *FEBS Lett.* **1987**, *213*, 265-268.
26. Fauchere, J.-L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369-375.
27. Katayama, T.; Nakao, K.; Akamatsu, M.; Ueno, T.; Fujita, T. *J. Pharm. Sci.* **1994**, *83*, 1357-1362.

RECEIVED May 1, 1995

Chapter 18

Successful Application of the QSAR Paradigm in Discovery Programs

Ernest L. Plummer

**Agricultural Chemicals Group, FMC Corporation, Box 8,
Princeton, NJ 08543**

The Quantitative Structure-Activity Relationship (QSAR) approach, first described by Corwin Hansch and Toshio Fujita thirty years ago, remains central to pesticide design. The successful application of this paradigm in the discovery of new pesticides depends on two key elements: statistical experimental design and effective computational tools. These tools must be available to bench scientists in a form that they can readily apply in the course of their daily design efforts. This report will describe a system that laboratory scientists have used for over ten years in our laboratory. Key components of this system are:

- Database tools to automatically integrate physicochemical parameters and biological data
- Experimental design tools for application of statistical experimental designs
- Analysis tools for pattern recognition and Free-Wilson and regression analyses

The report will illustrate the application of the approach to actual research programs. This will include application of cluster analysis, sequential simplex, and central composite designs to rapidly advance pesticide design projects.

Thirty years ago Drs. Corwin Hansch and Toshio Fujita initiated a body of work that has evolved into the current QSAR paradigm[1]. This represented a dramatic shift from the way drugs and pesticides had been discovered; it moved from an intuitive evaluation of the effects of structure change on biological activity to a quantitative, scientifically verifiable evaluation based on accurately measured physical properties and sound statistical principles.

Over the years that I have been involved in following and teaching the QSAR Paradigm, the question most asked has been: "Does QSAR really work?" It has become easier for me to answer the question as I see the many ways that QSAR can make Discovery programs more effective. Today I would list QSAR's success criteria as the following:

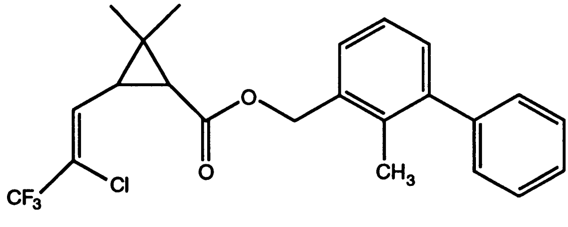
0097–6156/95/0606–0240$12.00/0

- Does it significantly increase the productivity of Discovery Programs? That is, does it provide evidence that the "best" compound has been found and do so at a minimum cost?
- Does it provide new leads?
- Does it provide commercial compounds?

Based on my personal experiences, I would answer all these questions, yes!

Examples of the Successful Application of the QSAR Paradigm.

The Discovery of a Commercial Compound. Very early in our application of the QSAR paradigm we experienced it's strength in rapid, cost effective optimization of a lead in a project that yielded what is now a commercial product, bifenthrin [2]. Bifenthrin was the product of a design strategy based on cluster analysis as originally introduced to drug design by Corwin Hansch and his coworkers [3]. The initial design that lead to the new biphenyl lead, required the synthesis of 18 compounds. These compounds were sufficient to develop a good model for structure activity relationships in the meta position of benzyl alcohols used to make pyrethroid esters. Optimization of the new biphenyl alcohol lead required about 40 additional compounds.



bifenthrin

The QSAR paradigm can successfully guide Discovery programs that lead to commercial products. But for every commercial compound there are a large number of compounds that you can optimize using QSAR strategies but, because of some shortcoming in performance or cost, do not ultimately qualify for commercialization. A major need in Discovery is a process to discover new leads and optimize them rapidly and efficiently. When this process is in place many more leads can be explored with the same resources required for traditional art-based design projects. But first one needs leads.

New Lead Generation. So can QSAR strategies generate new leads? New lead generation can take many forms:

- The ultimate goal is to create a model of existing compounds that will lead to a totally new lead. I have not experienced that success.
- A second approach is to use QSAR to develop models from which novel analogs of existing compounds can be developed. Bifenthrin was produced from working on an already commercial molecule. Success in this area is a very common occurrence.
- Many Discovery programs are using biological screening of thousands of randomly chosen compounds as a means of lead generation, i.e., massive screening. One may expect hundreds of active compounds to be detected by these programs. Many compounds show very low activity in these screens. Some investigators will only investigate compounds that reach some relatively high level of activity. However, I

believe that QSAR approaches offer a way to take even very low activity compounds to commercial levels.

Several years ago I was involved in a team effort to evaluate screening of randomly chosen compounds as a means of generating new leads. During the project we screened 6000 to 7000 randomly chosen compounds. Many compounds showed activity in the screen. We used QSAR strategies to evaluate over 60 leads of significantly different chemistry and mode-of-action. We found that activity could be optimized as much as three to five orders of magnitude simply by changing substituents. Thus, a compound that could barely meet the activity criteria of the screen could, on optimization, lead to a compound with a commercial level of activity. We also discovered that many compounds could not be improved over the levels of activity found in the initial lead. QSAR Strategies can determine this rapidly.

We learned that finding new leads from a program of screening randomly chosen compounds requires a systematic approach. Many discovery scientists depend on a set of experiential rules to optimize leads. These are not as successful when the mode- or mechanism-of-action is not known since there is no way to "map" knowledge-based rules onto the lead.

In a massive screening program one must evaluate hundreds of leads. The shear number of actives that are generated in a massive screening effort preclude exhaustive synthesis of analogs. QSAR based Statistical Experimental Designs offer an approach that can make these programs very efficient.

Efficient Optimization of Lead Molecules. Statistical Experimental Designs can be categorized roughly into five groups:

- Art-based designs
  - The Topliss Tree was one formalization of this approach [4].
- Univariate designs
  - The most fundamentally practiced design in science, that is, hold everything constant except one variable. This approach is easy to conceptualize but, like Art-based designs, is not very efficient.
- Random designs
  - Particularly stratified designs like Craig plots and the Cluster Analysis designs suggested by Corwin Hansch are very effective. Our own experience in the bifenthrin discovery program demonstrated that.
- Steepest Ascent Designs
  - Sequential Simplex Optimization
  - Factorial and Fractional Factorial Designs.
- Complete Factorial Designs
  - Central Composite Designs.

The last three designs, Random, Steepest Ascent, and Complete Factorial designs, are all efficient and rapid. But in recent years we have found that the Simplex Steepest Ascent Design and Central Composite Design are the best when either optimization of activity or development of a model for activity are the goal.

A Sequential Simplex Optimization. Simplex Designs are based on hill climbing techniques (Figure 1) introduced to Drug Design by Ferenc Darvas 20 years ago [5,6,7]. Here the first compound (Point **1**) is the least active compound in a three compound set. The simplex reflects from the least active compound through the center of a line connecting the most active compound (Point **3**) and the second most active (Point **2**). The line goes an equal distance beyond to the new compound (Point **4**). This compound is made and the process continues until the "top of the hill" or the most active compound is made. Sequential Simplex Designs are particularly useful in dealing with lead activity from massive screening since only n + 1 compounds, where n = the number of physical parameters being considered, need be prepared in the initial set. You require no pre-knowledge of what parameters are important to activity, since excess parameters will not cause the optimization to stray. We often consider sets like π, F, R, and the STERIMOL [8] parameters L and B_1 in an initial set of six compounds. One can expect optimization in six to ten additional compounds. For massive screening programs one can make the initial set, send it for evaluation, and then when all data are returned do a one-by-one follow-up. Reproducibility of data is less significant a problem as long as the most active, the second most active, and least active in each set can be identified; intermediate values, if present, are irrelevant to any given simplex. We used this approach on most of the compounds from our screening project with great success.

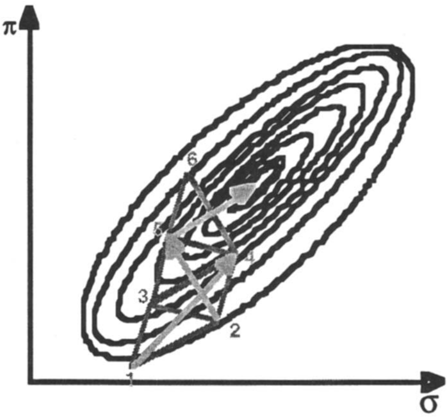


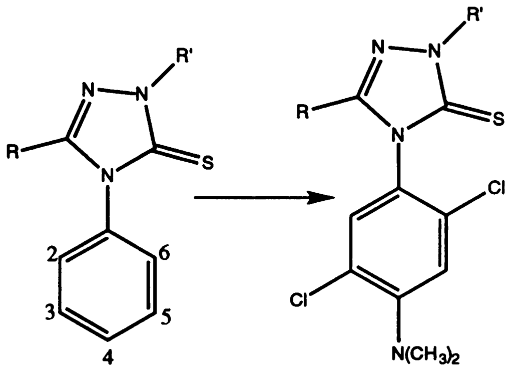
Figure 1. A Seven Compound Sequential Simplex Optimization in $\pi - \sigma$ Space

Recently we published an example of a herbicide optimization project [9]. The lead molecule (1) contained 7 non-equivalent positions. Our strategy was to use sequential simplex as the primary optimization tool.

Our first step was to use chlorine substitution of the aromatic ring as a probe to determine which positions were most sensitive to substitution [10]. Those positions that showed dramatic loss of activity were avoided. We concentrated our efforts on the positions where activity was increased or unaffected.

It turned out that the pattern that emerged from the Free-Wilson analysis of the chlorine probe set was 2,4,5-trisubstitution. We proceeded to optimize each position with a Simplex set. In the end, we found our best compound after the synthesis of 89 compounds. Considering that the average project of this type runs from 500-1000 compounds this is a very efficient approach. The overall activity increase for our laboratory evaluation was 25,000 fold. Although this compound came very close to commercial field activity, it was like many that get to the field but are not quite good enough.

A Central Composite Design. Simplex Designs are particularly effective for leads with very low activity. For leads with very interesting activity, we have found factorial designs to be very effective. Design is achieved by first calculating points that represent high and low values of each parameter. Statistical measures of high and low are used. For example you take all the substituents you want to choose from and



1

determine what the statistical distribution of the properties are. You then choose a point that is either one standard deviation above or below the mean for each parameter. For example, in Figure 2a, the point (-, -, -) is one standard deviation below the mean for π, σ, and mr, respectively. You then place these points in the three dimensional space represented by your set of substituents and choose substituents close to the points. When you're done you have eight substituents each representing two levels of each physicochemical parameter.

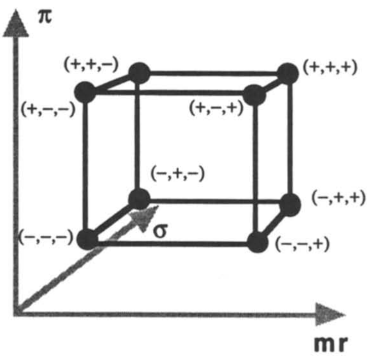


Figure 2a. 2^n Factorial Design; Two levels; 8 compounds.

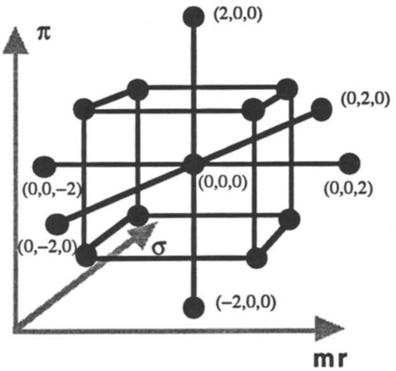
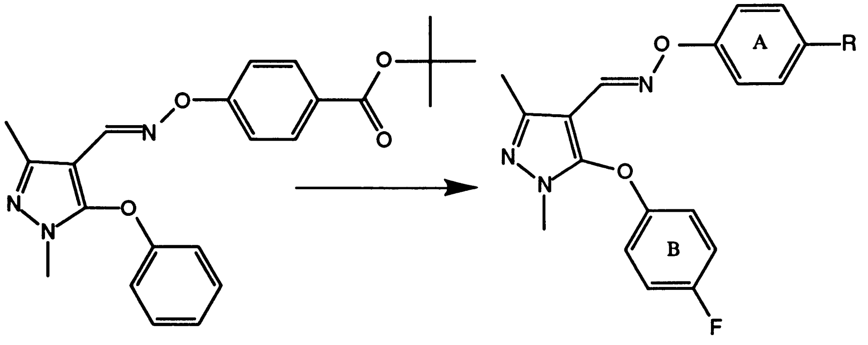


Figure 2b. Central Composite Design; Five Levels; 15 compounds.

A set like this is not large enough, statistically, to build a model. Recently [11], I suggested that a better design set could be obtained by using a central composite design. This is obtained by adding a star design to the factorial design. The star design has a center point at the mean for all variables and six points two standard deviations away, plus or minus, for each variable. The set of 15 compounds represents for three parameters five levels of each parameter. An ideal substituent set can easily be found using this design. The following is, to my knowledge, the first real application of a central composite design in drug or pesticide discovery.

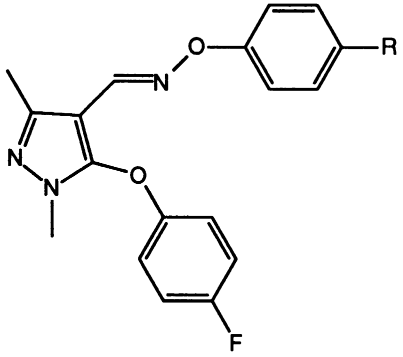
Recently we wanted to quickly see the potential of a novel insecticide, fenpyroximate (NNI 850), first prepared by scientists at Nihon Nohyaku [12]. We chose the central composite design to explore the area.

As is often our practice, we dealt with one ring at a time in Compound **2**. To minimize metabolic detoxification, by aryl hydroxylation, we blocked the para position of the B-ring with a fluorine atom. We used a chlorine probe strategy to determine that the para position on the A-ring was the most likely position for substitution; all other positions lost activity on substitution.



Fenpyroximate **2**

Using a database of 139 well characterized substituents [13] we created the design shown in Figure 3. We chose π, σ and mr (mr is used here to denote the molar refractivity substituent constant devided by 10) as our physicochemical parameters because they represent the three principle physical properties and provide a conservative set for synthesis. The fifteen substituents represent five levels of each substituents covering four standard deviations of the parameter space.



Substituent	Design Point (π, σ, mr)
$CH=CHCF_3$(c)	0, 0, 0
C_2H_5	+, -, -
CN	-, +, -
CF_3	+, +, -
$NHC=O(NHC_2H_5)$	-, -, +
OC_5H_{11}	+, -, +
$P=O(OC_2H_5)_2$	-, +, +
$C=O(OC(CH_3)_3)$	+, +, +
OCH_3	-, -, -
SCF_2CF_3	2, 0, 0
$C=ONHCH_3$	-2, 0, 0
SO_2CF_3	0, 2, 0
$OCH(CH_3)_2$	0, -2, 0
$OC=O(C_6H_5)$	0, 0, 2
F	0, 0, -2

Figure 3. A Central Composite Design in Three Parameters.

When you check the actual substituent parameters shown in Table I, you will recognize that the design does not conform exactly to the points in space depicted in Figure 2b for the central composite design. That is, of course, because the 139 compounds represent discrete points in a discontinuous space.

Our design program calculates the ideal addresses for each design point and then searches for compounds near the points. We limit the search so that the design does not become so over distorted as to cause a lack of orthogonality in the design. A list of substituents is presented to the chemist so that a choice is available to address synthetic feasibility. If the search fails to find acceptable substituents at one or more points one can rotate the design or change its size.

Table I provides the physicochemical values for the substituents in the design set.

Table I. Physicochemical Parameter Values

Substituent	Design Point	π	σ	mr
CH=CHCF$_3$(c)	0, 0, 0	0.66	0.17	1.56
C_2H_5	+, -, -	1.02	-0.15	1.03
CN	-, +, -	-0.57	0.66	0.63
CF_3	+, +, -	0.88	0.54	0.50
NHC=O(NHC$_2$H$_5$)	-, -, +	-0.50	-0.26	2.32
OC_5H_{11}	+, -, +	2.04	-0.34	2.63
P=O(OC$_2$H$_5$)$_2$	-, +, +	-0.53	0.60	3.12
C=O(OC(CH$_3$)$_3$)	+, +, +	1.21	0.45	2.61
OCH_3	-, -, -	-0.02	-0.27	0.79
SCF_2CF_3	2, 0, 0	3.41	0.48	1.91
C=ONHCH$_3$	-2, 0, 0	-1.27	0.36	1.46
SO_2CF_3	0, 2, 0	0.55	0.93	1.29
OCH(CH$_3$)$_2$	0, -2, 0	1.05	-0.45	1.71
OC=O(C$_6$H$_5$)	0, 0, 2	1.46	0.13	3.23
F	0, 0, -2	0.14	0.06	0.09

If one does not stray from the rules of selection, a set like this should fit the description of an "ideal" substituent set. An ideal substituent set can be defined as one in which:

- Parameters vary independently (orthogonal).
- Parameters are well spread in parameter space.
- The set is large enough for statistical significance (4 - 6 substituents per parameter).
- The set is metabolically stable.
- The set is synthetically feasible.

Do the parameters vary independently? Yes. Based on the correlation matrix and Factor analysis all parameters are orthogonal as shown in Table II; including π and mr which are often difficult to separate. Are the parameters well spread in space? Yes. The four standard deviations or five levels provide a large separation in the space as indicated by the parameter ranges in Table II.

Are there sufficient observations? Yes. For three parameters there will be five observations per parameter. Are they metabolically stable? Probably. None are extremely labile. Are they synthetically feasible? This is always arguable since synthetic feasibility is in the eye and the skill of the beholder. Since we were able to

Table II. Correlation Matrix, Factor Loadings and Parameter Ranges

	Correlation Matrix			Factor Loadings			Parameter Ranges
	π	σ	mr	Factor 1	Factor 2	Factor 3	
π	1.000			0.998	0.000	0.000	-1.27 ~ 3.41
σ	-0.101	1.000		0.000	0.992	0.000	-0.45 ~ 0.93
mr	0.247	-0.070	1.000	0.000	0.000	0.991	0.09 ~ 3.23

prepare all but one of these compounds I think the answer here is yes. Thus the central composite design was able to select an ideal set. We find this to be generally true.

The one exception to synthetic feasibility was the design's center point, the trifluoromethyl vinyl substituent. When we found that we were unable to make this substituent, we searched the space near it and found that the Iodo group could be substituted without destroying either the orthogonality or the spatial representation of the set. The fifteen compounds were synthesized and then tested against the larvae of yellow fever mosquito (Table III).

Table III. Physicochemical Parameter Values and Biological Activity

Substituent	Design Point	π	σ	mr	YFM* pI_{50}
$CH{=}CHCF_3$(c)[I]	0, 0, 0	0.66 [1.12]	0.17 [0.18]	1.56 [1.39]	6.2
C_2H_5	+, -, -	1.02	-0.15	1.03	6.1
CN	-, +, -	-0.57	0.66	0.63	5.8
CF_3	+, +, -	0.88	0.54	0.50	6.6
$NHC{=}O(NHC_2H_5)$	-, -, +	-0.5	-0.26	2.32	nm
OC_5H_{11}	+, -, +	2.04	-0.34	2.63	4.5
$P{=}O(OC_2H_5)_2$	-, +, +	-0.53	0.60	3.12	4.5
$C{=}O(OC(CH_3)_3)$	+, +, +	1.21	0.45	2.61	6.4
OCH_3	-, -, -	-0.02	-0.27	0.79	5.8
SCF_2CF_3	2, 0, 0	3.41	0.48	1.91	7.3
$C{=}ONHCH_3$	-2, 0, 0	-1.27	0.36	1.46	4.9
SO_2CF_3	0, 2, 0	0.55	0.93	1.29	7.5
$OCH(CH_3)_2$	0, -2, 0	1.05	-0.45	1.71	5.8
$OC{=}O(C_6H_5)$	0, 0, 2	1.46	0.13	3.23	4.0
F	0, 0, -2	0.14	0.06	0.09	5.4

*YFM = yellow fever mosquito larvae

We decided to use the yellow fever mosquito assay since the assay was rapid and very stable, producing data of low variability. You will note that one substituent, the ethylureido group produced a compound that was not active at the limits of the test. This is often the case when one uses an ideal design since it spreads the compounds in activity space. The loss of this compound did not destroy the ideal nature of the set, allowing us to continue with our analysis. Instead of 15 compounds we had 14; more than four observations per parameter.

Our general approach to analysis is to first run and examine plots of each independent variable against the dependent or activity variable. We do this to observe any clustering or highly influential points and to observe any curvature in the plots that might indicate an optimum in any of the variables. We do this instead of adding all squared terms; a process that could lead to chance correlation.

We then run all regressions for the independent variables and combinations of independent variables up to the limit of four observations per independent variable. In our computer system this is under algorithmic control. Statistically the best model was a three component model in π, σ and mr (equation 1). Note that only the design parameters were tested when constructing the model, reducing the probability of chance correlation. Sixty-four percent of the biological variation can be explained by the three parameters, an acceptable result for whole organism data.

$$pI_{50}\,(\mathrm{YFM}) = 0.46\,\pi + 1.34\,\sigma - 0.56\,\mathrm{mr} + 6.06 \qquad (1)$$

$n = 14$ $r = 0.797$ $r^2 = 0.635$ $s = 0.680$ $F = 5.79\ (p= 0.015)$

π: $t = 2.71\ (p = 0.02)$, σ: $t = 2.89\ (p = 0.02)$, mr: $t = -2.79\ (p = 0.02)$

When we examined our plots of activity *vs.* each physicochemical property, we noted curvature in the mr *vs.* activity plot suggesting an optimum in that parameter. We ran the squared mr term equation 2. Although there are only a bit over three observations per independent variable, this appears to be a statistically sound equation. The optimum value for mr would be 1.14. Nearly 80% of the biological variation is explained by the equation.

$$pI_{50}\,(\mathrm{YFM}) = 0.38\,\pi + 1.35\,\sigma + 1.03\,\mathrm{mr} - 0.45\,\mathrm{mr}^2 + 5.14 \qquad (2)$$

$n = 14$ $r = 0.886$ $r^2 = 0.785$ $s = 0.550$ $F = 8.21\ (p = 0.004)$

π: $t = 2.65\ (p = 0.03)$, σ: $t = 3.60\ (p = 0.01)$, mr: $t = 1.58\ (p = 0.15)$,
mr^2: $t = -2.51\ (p = 0.03)$, $\mathrm{mr}_0 = 1.14$

In the process of doing the chlorine probe set, we made the unsubstituted and the chloro substituted examples (Table IV). When these are predicted by the equation along with the one unmeasurable compound in the design set, we got acceptable fit.

The logical move was to add the two compound to the original set. When we did that we got equation 3.

$$pI_{50}\,(\mathrm{YFM}) = 0.37\,\pi + 1.33\,\sigma + 1.04\,\mathrm{mr} - 0.45\,\mathrm{mr}^2 + 5.10 \qquad (3)$$

$n = 16$ $r = 0.882$ $r^2 = 0.779$ $s = 0.509$ $F = 9.07\ (p = 0.001)$

π: $t = 2.82 (p = 0.02)$, σ: $t = 3.89\ (p = 0.00)$, mr: $t = 2.01\ (p = 0.07)$,
mr^2: $t = -3.02\ (p = 0.03)$, mr_0: $= 1.16$

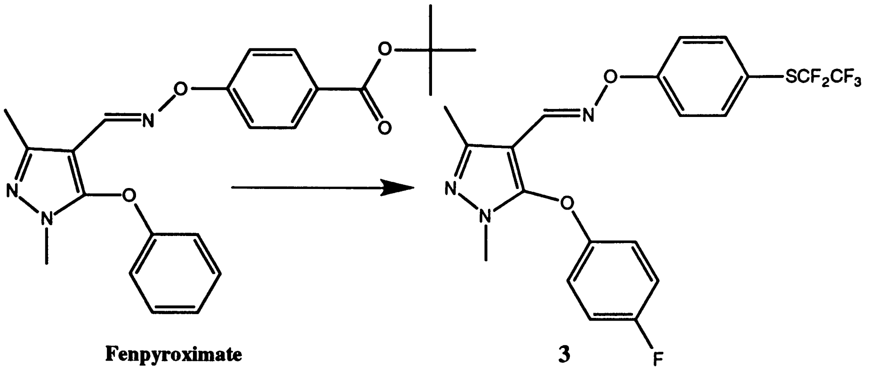
When the two new compounds were added to the design set, bringing the observations to four per independent variable, the equations were essentially the same. Our conclusion is that using a central composite design we not only could easily create an ideal substituent set but that this set, when tested, produced a model that was reasonably rigorous and explained the biological variation well.

Table IV. Additional Substituents

Substituent	π	σ	mr	pI_{50} (YFM) Predicted	pI_{50} (YFM) Observed
H	0.00	0.00	0.10	5.3	5.3
Cl	0.71	0.23	0.60	6.2	5.8
$NHC{=}O(NHC_2H_5)$	-0.50	-0.26	2.32	4.6	nm

At the end of our analysis we used our model and QSAR system tools to examine our full physicochemical database. This larger database contains substituents for which some of the parameters have been estimated or calculated. When we did so we found some other substituents that might produce equal activity but none, within the design space, that might be better.

We really have no reason to develop a mosquito control agent. Our target is agricultural pests. But the same well designed set can be used to directly test in the somewhat less controlled environment of foliar assays. In Figure 4, you can see that we obtained the Compound **3** with good worm and mite activity without the preparation of additional compounds. A clear benefit of a well designed initial set.



Fenpyroximate

3

Insect	Activity
pI_{50} (YFM)	nm
LC_{50} (TBW)	930 ppm
LC_{50} (TSM)	~2.0 ppm

Insect	Activity
pI_{50} (YFM)	7.3
LC_{50} (TBW)	87 ppm
LC_{50} (TSM)	5.0 ppm

TBW: tobacco budworm
TSM: two-spotted spider mite

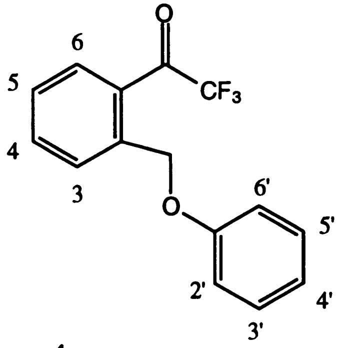
Figure 4. Foliar Activity Against Other Pests

In dealing with leads generated from the screening of randomly selected compounds we discovered two things:

- The application of statistical design strategies in a systematic optimization process was the most effective way to develop leads to their full potential, but...

- One cannot optimize all leads. Some are at or very near their maximum potential when first discovered.

The trifluoromethylketone based herbicide 4, is an example of a lead that could not be optimized. As indicated it had nine non-equivalent aryl positions. We applied a strategy that involved chlorine and fluorine probe atoms in each position and a series of sequential simplex sets [7].



4

Although we applied our best approach, even giving extra effort by looking at several positions and combined positions, we were unable to optimize even after the expenditure of resources for the synthesis of 54 compounds. However, our systematic approach allowed us to abandon this lead with the full confidence that we had seen the best in the series.

The third criteria of success was efficient optimization. I think I have demonstrated with these examples that here too the QSAR Strategies are successful. We have repeated similar studies with equal success.

Does QSAR "work"? Yes! Then why doesn't it work all the time? Like I just indicated some leads can't be optimized. But sometimes its because it isn't done right. I suggest that many of the projects that don't work, fail because the compounds chosen for synthesis were not chosen by an effective design strategy.

But if it works so well, why doesn't everyone use QSAR approaches. In my own experience the availability of computational tools is the single most important factor followed by training to use those tools.

Managing QSAR Projects

Even when tools are available to effectively implement QSAR strategies some problems arise if the wrong people are asked to use the programs. The problem is to get the programs used by the right people. The right person is the laboratory chemist – the design scientist.

The laboratory chemist is the individual with the relevant information for making design decisions. She or he has a series of skills or knowledge-bases unavailable to the average Computer-aided molecular design expert. These include knowledge of methods of synthesis, metabolic fate, compound-specific knowledge, and domain (*e.g.*, herbicide or insecticide) specific knowledge. When following a lead in an area where considerable knowledge exists this becomes extremely important.

On the other hand, the computer-aided molecular design expert has skills and knowledge difficult to find in the average laboratory chemist. These include statistical experimental design, statistics, quantum and molecular mechanics, and knowledge of computer software and hardware.

The key is to put the tools to execute computer-aided molecular design into the hands of the laboratory chemist. The approach may be somewhat different depending on the tool:

- Molecular modeling tools require extensive specialized knowledge. The approach we follow is to get the design specialist working hand-in-hand with the laboratory scientist.
- QSAR does not require the same level of specialized knowledge. The tools should be in the hand of the laboratory scientist.

The trick is to get the design expert and the laboratory scientists working together as a team but not doing the same things.

A Model for Design Project Management. We view the laboratory chemist as the customer of the computer-aided molecular design team. The design expert is there to prepare and maintain computer tools that the chemist can use or to act as a facilitator for the chemists use of complex tools. Their job includes conducting training. The centerpiece of our tools is a QSAR system that we have developed over the course of fifteen years.

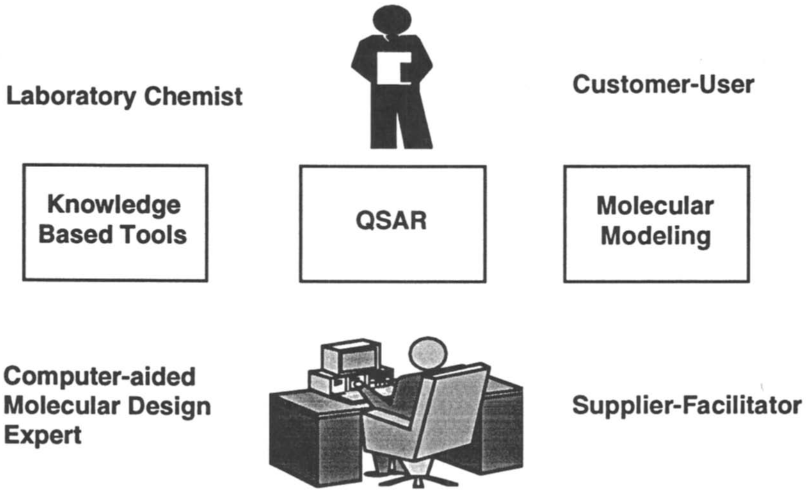


Figure 5. Design Project Management Model

System Components. A necessary but not sufficient condition for success in the QSAR arena is the construction of a QSAR on-line system. In our system we have incorporated three elements (Figure 6):

- Design Tools, because we think they are of paramount importance.
- Database Tools because assembling data on sets is very time consuming and error prone. A process that we see as a significant barrier to applying QSAR strategies.
- Analysis Tools.

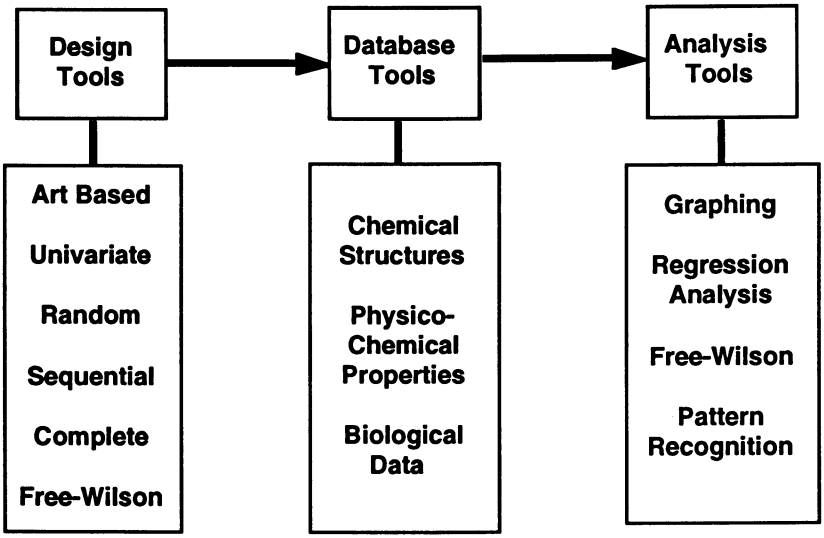

Design Tools
Database Tools
Analysis Tools
Art Based
Univariate
Random
Sequential
Complete
Free-Wilson
Chemical Structures
Physico-Chemical Properties
Biological Data
Graphing
Regression Analysis
Free-Wilson
Pattern Recognition

Figure 6. Outline of QSAR System Modules and Components

To make these tools truly available to the laboratory chemist it is necessary for them to be user-friendly. A uniform, unified system being the key to lowering the barrier to acceptance. Spreadsheet like presentations also helps in this area.

We have found that the physicochemical properties available in the literature for substituents are limiting. Only 300 to 400 substituents have complete data. This set does not represent a large enough or uniform enough physicochemical space to run design strategies like sequential simplex effectively. We therefore have used estimation programs to fill in gaps in published data on substituents and invented new substituents to enlarge the space.

Conclusion

The work begun 30 years ago by Dr. Fujita and Dr. Hansch has proven to be a very effective tool for managing our Discovery programs. The key to making it work is putting the tools in the appropriate hands and applying effective statistical experimental design strategies as the starting point.

Acknowledgments

The author acknowledges the significant contributions of the following scientists to this work; for synthesis: Brian Bennett, Alex Oliphant, and Andrew Baver, for biological testing: Mike Walsh, Joan Plummer, and Donata McLean, and for analytical support: Bill Creekmore.

Literature Cited

1. Hansch, C.; Fujita, T. **1964**, *86*, 1616 - 1625.
2. Plummer, E. L. In *Pesticide Synthesis Through Rational Approaches*; P. S. Magee, G. K. Kohn and J. J. Menn, Ed.; American Chemical Society: Washington, 1984; Vol. 255; pp 295 - 320.
3. Hansch, C.; Unger, S. H.; Forsythe, A. B. *J. Med. Chem.* **1973**, *16*, 1217.

4. Topliss, J. G. *J. Med. Chem.* **1972**, *15*, 1006.
5. Darvas, F. *J. Med. Chem.* **1974**, *17*, 799 - 804.
6. Gillom, R. D.; Purcell, W. P.; Bosin, T. R. *Eur. J. Med. Chem.* **1977**, *12*, 187.
7. Plummer, E. L. In *Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals*; W. Draber and T. Fujita, Ed.; CRC Press: Boca Raton, 1992; pp 3-39.
8. Verloop, A.; Hoogenstraaten, H.; Tipker, J. In *Drug Design*; E. Ariens, Ed.; Academic Press: New York, 1976; Vol. 7; pp 165 - 207.
9. Simmons, K. A.; Dixson, J. A.; Halling, B. P.; Plummer, E. L.; Plummer, M. J.; Tymonko, J. M.; Schmidt, R. J.; Wyle, M. J.; Webster, C. A.; Baver, W. A.; Witkowski, D. A.; Peters, G. R.; Gravelle, W. D. *J. Agri. Food Chem.* **1992**, *40*, 297-305.
10. Plummer, E. L. In *Reviews in Computational Chemistry*; First ed.; K. B. Lipkowitz and D. B. Boyd, Ed.; VCH Publishers, Inc.: New York, 1990; pp 119 - 168.
11. Plummer, E. L. In *Advances in International Research, Development, and Legislation*; VCH: Hamburg Germany, 1990; pp 51-60.
12. Hamaguchi, H.; Ohshima, T.; Konno, T.; Miyagi, Y.; Shiraiwa, Y.; Akita, T.; Nihon Noyaku Co., Ltd.: U.S. 4,843,068, June 27, 1989.
13. Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979.

RECEIVED April 26, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

Chapter 19

Comparative QSAR: Understanding Hydrophobic Interactions

Corwin Hansch

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711–6338

In our current database of 3000 bio QSAR less than 10% lack hydrophobic terms. It is of particular interest that some of these are for QSAR in whole organisms. Various examples are given. A surprising number involve electrophiles. Also in instances where reaction has been shown to occur with a receptor without apparent hydrophobic interactions, the same effect is present in the QSAR for the cell or organism.

Since the systematic development of octanol/water partitioning as a model for hydrophobic interactions began over 30 years ago (*1*), logP and π have been important in the derivation of thousands of quantitative structure-activity relationships (QSAR). Clearly the methodology has proved to be valuable in the design of bioactive and less toxic chemicals (*2*). Nevertheless, the field has grown in a haphazard manner with most workers doing little to show how their new QSAR fits in with what has been done. In searching for more order we have developed an easily searchable computerized database of about 6,000 QSAR about evenly divided between the purely chemical and biological fields (*3,4*). It is hoped that this report will illustrate how such a database can be used by focusing on QSAR which do not contain a hydrophobic term or which contain a small negative term.

A search of our present database, except those for macromolecules and isolated enzymes where membrane penetration or interaction with hydrophobic cell components is not possible, finds 1918 QSAR. Eliminating those with hydrophobic terms yields 143 equations (7.5%). In a number of these it is likely that hydrophobic terms have been missed because of poor experimental design resulting in collinearity between steric and hydrophobic properties. Thus, we see that it is rather unusual to find examples where hydrophobic interactions can be neglected. Finding such a QSAR means that extra effort should be made to understand the meaning behind it.

Organic compounds can undergo two types of hydrophobic interactions: 1) specific interaction with a hydrophobic part of a receptor. 2) More or less nonspecific interactions with macromolecules (eg. serum proteins) and membranes. Little has been done to delineate these two processes. It is generally assumed that small molecules (MW < 500) penetrate membranes by hydrophobically facilitated passive diffusion; however, size can be very critical (*5*).

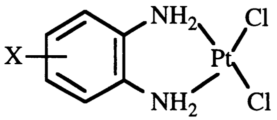
0097–6156/95/0606–0254$12.00/0

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

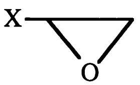
QSAR Where DNA is the Target

We first consider QSAR which likely correlate reactions with DNA. In a summary of QSAR for bacterial mutagenesis (*6*), 11 examples were found for aromatic amines and nitro compounds, nitrosoamines, nitrofurans and 2-aminobenzimidazoles acting on a variety of bacterial cells.

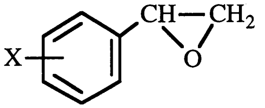
All of these compounds require activation by microsomal S9 or cytosolic reductase, and hydrophobic terms are essential. In ten of the 11 examples, the coefficient with the hydrophobic term is essentially 1. However, for the following four examples of direct acting electrophilic mutagens, hydrophobic terms were of no help, only electronic terms were essential (*6*).



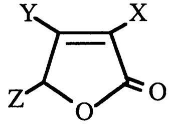
1



2



3



4

One must bear in mind that these studies were made on relatively small sets of compounds with relatively small substituents. Compounds with larger hydrophobic groups, say long alkyl units, might require hydrophobic terms.

The mutagenicity of the following two direct acting alkylating agents is related to this point.

R_1OSO_2Me
5

$R_1OSO_2R_2$
6

In the two QSAR for the above sulfonate esters it was essential to parameterize R_1 using π, but despite wide variation in the hydrophobicity of R_2 no parameterization was needed for this part of the molecule (*6*). Congener 6 suggests that these compounds are able to cross the cell membrane and move through the cytoplasm to reach DNA without rate limiting hydrophobic interactions. The fact that R_2 shows no hydrophobic effect suggests membrane crossing is not hydrophobically significant. It would seem that the hydrophobic effect of R_1 must be the result of an interaction with the base pairs of the DNA.

We recently reviewed the QSAR of the aniline mustard antitumor agents: $X\text{-}C_6H_4N(CH_2CH_2Y)_2$; Y = halogen or OSO_2R (*7*). It is presumed that alkylation of tumor DNA is the source of antitumor activity. Four examples were found for comparison.

Leukemia
25% increase in life span of mice
$Y = OSO_2R$
$\log 1/C = -0.35 \log P - 1.31\,\sigma^- + 3.08\,I$
$n = 16, r = 0.970$
I

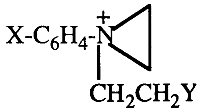
Leukemia
25% increase in life span of mice
$Y = Cl$
$\log 1/C = -0.31\,\pi - 0.96\,\sigma^- + 0.86\,I$
$n = 19, r = 0.926$
II

Melanoma
25% increase in life span of mice
$Y = Cl$
$\log 1/C = -0.42 \log P - 1.70\,\sigma$
$n = 20, r = 0.919$
III

Walker solid tumor
90% cure of rats
$Y = Cl, Br, I$
$\log 1/C = -1.52\,\sigma^- + 0.57\,I_1 + 0.44\,I_2$
$n = 17, r = 0.931$
IV

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

In example I, I = 1/0 for the presence or absence of 4-N=O. In example II, I = 1 for ortho substituents. In example IV, I_1 = 1 for Y = Br or Cl and 0 for I and I_2 = 1 for the presence of a COOH group. Again we can find no evidence for a hydrophobic term in this equation. In example III for the substituents studied, σ^- does not differ significantly from σ, hence σ appears in this QSAR. The average value of ρ (the coefficient with σ) is -1.37. The evidence suggests that an internal nucleophilic displacement of Y by N occurs to yield an aziridinium ion:



which reacts electrophilically with DNA. One QSAR contains no hydrophobic term and three contain small negative terms. Another study of the mustards producing an LD_{50} within 21 days in rats yields a QSAR (*7*): log 1/C = $-1.38\,\sigma^- + 0.41\,I$, n = 17, r = 0.930. I = 1 for Y = Br and 0 for Y = Cl or I.

The above results can be compared with $X\text{-}C_6H_4SCH_2CH_2Br$ producing an LD_{50} with red spider eggs (*7*): log 1/C = $-1.42\,\sigma^- + 1.49\,I$, n = 14, r = 0.956, I = 1 when X = COOR. These compounds also appear to form onium intermediates which then produce the toxic interaction. Thus, ρ in QSAR for toxicity to eggs and LD_{50} for rats is in agreement with average value for the four equations for antitumor activity. In all of these examples, hydrophobic terms are missing or are negative. Five of the six examples are for whole animals, where one would most likely expect to find the hydrophobic properties of chemicals to be important. There is every opportunity for hydrophobic interactions with serum proteins or membranes in the random walk of chemicals to their reaction sites.

Isothiocyanates

Another class of compounds which reacts with nucleophiles is: $X\text{-}C_6H_4N{=}C{=}S$. Their toxicity (I_{50}) to *E. coli* cells is well correlated by σ alone, ρ = 2.27. The addition of hydrophobic terms does not improve the correlation (*3*). It is of interest that the toxicity of the much less electrophilic R-N=C=S (R=alkyl) analogs is highly dependent only on logP (*3*). Two quite different mechanisms must be involved, but only the electrophilic isothiocyanates operate without a hydrophobic effect. Evidence that the phenyl isocyanates do react with nucleophiles in their toxic action to *E. coli* comes from the study of their reaction with ethanol (ρ = 2.16) and aniline (ρ = 2.14). These values of ρ are surprisingly close to 2.27 for their reaction with *E. coli.*

Folate Synthetase Inhibitors

One of the earliest QSAR was found in 1941 by Bell and Roblin who showed that the antibacterial activity of sulfonamides when plotted against their pKa values gave a parabola. Actually using a bilinear model gives a slightly better fit accounting for 87% of the variance in the data. That is, a rather good correlation is obtained without a hydrophobic term. Seydel (*8*) followed up this study using $H_2N\text{-}C_6H_4SO_2NHC_6H_4\text{-}X$ acting on *M. smegmatis.* He obtained a good correlation with pKa alone (n = 36, r = 0.962). Addition of a hydrophobic term provides no improvement. In another study (*9*), using the same compounds with *E. coli* he found a correlation with σ, which we find can be improved slightly by the addition of a term in CLOGP.

$$\log 1/C = 1.32\,(\pm.23)\,\sigma - 0.24\,(\pm.14)\,\mathrm{CLOGP} + 5.28 \qquad (1)$$
$$n = 16,\ r = 0.963,\ s = 0.127$$

It is of particular interest that the coefficient with CLOGP (calculated logP (*10*)) is negative and small. In fact, it is reminiscent of the examples discussed above with the antitumor agents. Thus, there may be a small negative hydrophobic effect.

Seydel and his colleagues followed up this study using the same compounds to inhibit folate synthetase (*11*). (It has been well established that sulfonamides inhibit bacterial growth by inhibition of this enzyme). They found a good correlation with pKa which cannot be improved by the addition of a hydrophobic term (-0.32 pKa, n = 14, r = 0.930). This establishes the polar character of the enzymic active site. Thus, it would appear that the primary determinant of the QSAR is the nature of the receptor and that lipophilic membranes or the cell plasma have, at best, a small independent role (<10%). With many of the sulfa drug studies the QSAR are somewhat compromised by the collinearity between hydrophobicity and ionization. The apparent logP of partially ionized compounds is quite different from that of the neutral compound and this must be taken into consideration. Tied to this is the question of which form of the drug (ionized or unionized) is the active species.

Cholinesterase Inhibitors

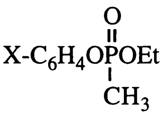
We now consider the inhibition of fly head cholinesterase *in vitro* and *in vivo* where possible hydrophobic interactions can be compared (*12*) as with the folate synthetase inhibitors. The following examples are for the inhibition of the fly enzyme.

$X\text{-}C_6H_4OP(OEt)_2$ (*13*)

$2.66\,\sigma^- - 0.47\,E_s\text{-}3$

$n = 12, r = 0.976$

V

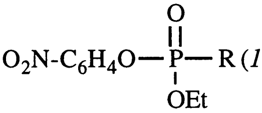


$X\text{-}C_6H_4O\underset{CH_3}{\overset{O}{\overset{\|}{P}}}OEt$ (*14*)

$2.94\,\sigma^- - 0.54\,E_s^c\text{-}3$

$n = 15, r = 0.946$

VI

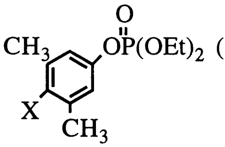


$O_2N\text{-}C_6H_4O{-}\underset{OEt}{\overset{O}{\overset{\|}{P}}}{-}R$ (*15*)

$2.71\,E_s^c$

$n = 16, r = 0.945$

VII

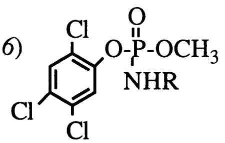


(*16*)

$2.62\,\sigma^- - 0.77B5$

$n = 10, r = 0.981$

VIII

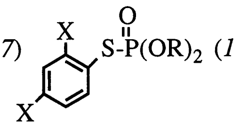


(*17*)

$1.25\,E_s^c$

$n = 7, r = 0.973$

IX



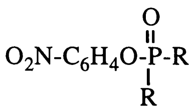
(*18*)

$1.63\,MR_4 + 1.44\,I$

$n = 12, r = 0.936$

$I = 1$ for $R = Et$ and 0 for $R = CH_2CHMe_2$

X

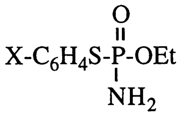


$O_2N\text{-}C_6H_4O\text{-}\underset{R}{\overset{O}{\overset{\|}{P}}}\text{-}R$

$1.28\,E_s^c - 1.65\,\sigma^*$ (*19*)

$n = 12, r = 0.950$

XI



$X\text{-}C_6H_4S\text{-}\underset{NH_2}{\overset{O}{\overset{\|}{P}}}\text{-}OEt$

$1.35\,\sigma^-$ (*20*)

$n = 8, r = 0.978$

XII

A rather complex QSAR for 269 carbamate inhibitors contains no hydrophobic terms (*21*).

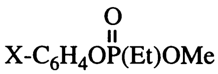
In the above examples the terms in the QSAR, the number of data points and correlation coefficients are listed under each structure. E_s is Taft's steric parameter

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

and E_s^c is Hancock's variation of E_s. Substituents in the meta position produce a positive steric effect (negative E_s terms) in examples V-VI. In none of the above examples do we see a hydrophobic term.

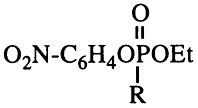
These results can be compared with activity (LD_{50}) in flies.



$X\text{-}C_6H_4OP(=O)(Et)OMe$

$2.87\ \sigma$ (*22*)
$n = 7, r = 0.977$
range in log P: 0.33 to 2.40

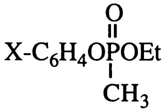
XIII



$O_2N\text{-}C_6H_4OP(=O)(R)OEt$

$-0.66 \log P$ (*23*)
$n = 10, r = 0.939$
range in log P: 1.6 to 4.0

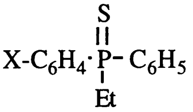
XIV



$X\text{-}C_6H_4OP(=O)(CH_3)OEt$ (*24*)

$-0.34 \log P + 1.84\ \sigma$
$n = 11, r = 0.962$
range in log P: 1.33 to 4.1

XV



$X\text{-}C_6H_4\cdot P(=S)(Et)\text{-}C_6H_5$

$2.13\ \sigma^- + 0.69\ MR\text{-}3$
$n = 11, r = 0.963$
range in log P: 4.3 to 6.1

XVI

Example XVI brings out the positive steric effect of meta substituents seen with the enzyme QSAR. In two of the examples there is no hydrophobic term and in the other two there is negative dependence of activity on hydrophobicity. In example XIV there is only a hydrophobic term, but it is negative. Thus examples I to XVI show that even in whole organisms there seems to be no positive hydrophobic effect. It is the character of the receptor which appears to determine the shape of the QSAR, and not membrane crossing or cytosolic material. Examples V to XII concern reactions with nucleophiles, as do examples I-IV.

Radical Toxicity of Phenols and Anilines

Many equations have been published on the toxicity of phenols to all kinds of organisms or their parts (DNA, enzymes, organelles, cells, membranes). Our present database contains over 100. In all of these, activity is a function of hydrophobicity and in a few cases a positive σ term. There are many similar examples for anilines. In stark contrast to this we have found (*7*) from a study by Karlock and his associates (*26*) for the toxicity of phenols to rat embryos *in vitro* a dependency on σ^+ ($\rho = -0.6$) alone. In a fourth case of phenols causing weight loss in pregnant rats a similar value of ρ was found. With the pregnant rats logP did play a role, however; in this example it is likely that a greater variety of toxic effects may be in play.

Somite test
$\log 1/C = -0.65\ \sigma^+, n = 11, r = 0.910$
XVII

Tail defect test
$\log 1/C = -0.58\ \sigma^+, n = 10, r = 0.912$
XVIII

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

Tail defect test with hepatocytes
$\log 1/C = -0.56\ \sigma^+,\ n = 9,\ r = 0.913$

XIX

Weight loss of pregnant rat
$\log 1/C = 0.83 \log P - 0.18 (\log P)^2 - 0.63\ \sigma^+$

$n = 19,\ r = 0.865$

XX

The addition of hydrophobic terms did not produce significant improvement in examples XVII to XIX (*7*). The unusual aspect of the study by Oglesby et al. (*26*) was that they made an excellent selection of substituents including a number with large negative σ^+ values (*26*). Using hydrophobic phenols with electron withdrawing substituents would have produced an entirely different result.

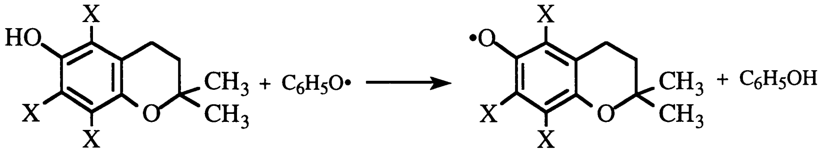
Example XX, in which the end point was rather drastic (10 g weight reduction in 24 hrs in pregnant rats), hydrophobic terms are important. This is associated with σ^+ which has the same ρ value as in examples XVII to XIX. Of course what all goes into rapid weight loss by a pregnant rat is not clear. It seems likely that much more than one mechanism is involved. In example XX the σ^+ term correlates a larger fraction of the variance than the two logP terms.

In a different type of study by Richard et al. (*27*) on the 50% inhibition of DNA synthesis in Chinese hamster V79 cells by phenols we have found from their data a similar correlation based on σ^+ alone: $-0.85\ \sigma^+$, $n = 9$, $r = 0.849$, $F_{1,7} = 18.1$. Although the correlation is not very high it is quite significant, $F_{1,7}\ \alpha{:}\ 0.01 = 12.2$. Two data points were omitted: 2-$NHCOCH_3$ and 2,6-di-CH_3-4-$NHCOCH_3$.

At this point we turn to studies on phenols by physical organic chemists in which σ^+ is crucial. One of the advantages of our QSAR database is that we can range search for all equations involving phenols containing a σ^+ term. Two examples are found with similar ρ values involving homolytic oxidations which depend on σ^+ alone (*28*).

$$X\text{-}C_6H_4OH + (CH_3)_3CO\bullet \longrightarrow X\text{-}C_6H_4O\bullet + (CH_3)_3COH$$

$-0.71\ \sigma^+,\ n = 7,\ r = 0.990$ XXI



$-1.11\ \sigma^+,\ n = 9,\ r = 0.910$[29]

XXII

This suggests that radical formation may be the cause of this type of toxicity.

The phenol toxicity can be compared with a report on the toxicity of anilines to Balb/3T3 embryo cells by Harada et al. (*30*) The study was based on an excellent selection of substituents in the 2,3 and 4-positions of the aniline ring. They correlated their data with five terms: logP, $(\log P)^2$, σ, HB1 and HB2; $n = 19$, $r = 0.887$. We have found a better correlation (including one more data point) with two terms: $-0.55\ \sigma^+ - 0.22 \log P$, $n = 20$, $r = 0.921$. More important; however, is the correspondence of σ^+ and its ρ value to those of the chemically similar phenols. The small negative logP is of importance since it is in line with our above discussion. We believe that this implies interaction with a nonhydrophobic receptor. That is, the phenols and anilines appear to be activated to radicals in an aqueous phase. The lack of hydrophobic terms implies that oxidation by P-450 enzymes is not likely since this generally depends strongly on hydrophobic interactions (*31*). The best candidate for the oxidant would seem to be •OH from the superoxide system. Both O_2^- and •OH are hydrophilic and would be retained in aqueous compartments. There are other reasons for suspecting the •OH as being the oxidant. It is extremely reactive and attacks almost every organic compound with which it comes in contact (*32*). The toxicity of radicals to DNA as well as other biological entities has become generally recognized (*33*). Its high reactivity means that the effect of the substituent is of rather low importance. The small values of ρ^+ for the phenol and aniline toxicity are in agreement with this point. Moreover, the very small •OH would likely not show steric effects with ortho substituents. In fact, none could be found for the toxicity of the anilines to fibroblast cells.

Discussion

As noted above, it is unusual to find QSAR without a positive hydrophobic term; hence it is important to consider what may lie behind such phenomena. Most easily documented is the case where the receptor does not present hydrophobic space with which hydrophobic parts of the ligand can make contact. This is clearly the situation with examples V to XII as well as the inhibition of folate synthase mentioned above. In contrast to inhibition of fly cholinesterase we have many equations for vertebrate enzyme which require positive hydrophobic terms. The X-ray crystallography of this enzyme shows it to contain a large hydrophobic active site (*34*). We have found a similar situation with the enzyme dihydrofolate reductase. In this instance the vertebrate enzyme reacts with inhibitors at three hydrophobic points while the bacterial enzyme makes only one hydrophobic contact. In these two examples the X-ray structures are available for comparison with the QSAR (*35*).

What is more interesting are examples where QSAR for the isolated enzyme can be compared with the enzyme in the organism. To reach the enzyme *in situ* inhibitors or substrates must cross a hydrophobic membrane in the simplest case of a bacterium. Despite this no hydrophobic term is seen for folate synthetase inhibitors. How can this be rationalized? It may be that the QSAR is determined largely by the nature of the receptor and that crossing a single membrane has a very small effect. Some bacteria contain channels (*36*) (porins) through which hydrophilic compounds may enter, and of course in certain instances active transport can solve the entry problem. The nature of the transport system may be either hydrophobic or hydrophilic.

The lack of a hydrophobic term for isothiocyanates might be explained by the above mechanisms, but with these compounds we have no idea of the receptor except that it is nucleophilic. It is possible that they might react with nucleophiles on the surface of bacteria. When the N=C=S moiety is insulated from the electrophilic phenyl ring in:
$X\text{-}C_6H_4CH_2N{=}C{=}S$, a quite different QSAR is found which contains only a logP term. These compounds are about 10 times less potent than their phenyl analogs.

Much more difficult to rationalize are the data on mutagenesis. So far in all of the examples (11) (*6*) where activation is required for mutagenesis a hydrophobic term appears in the QSAR. This is true whether activation occurs outside the bacteria or within the bacteria via reductases. The strong electrophiles 1 - 4 show no hydrophobic dependence. They remind us of the isothiocyanates. Maybe they too cause toxicity by reacting with a site on the membrane surface. This would have to be a nucleophile which does not present a hydrophobic patch. It is possible to imagine this occuring with DNA.

The QSAR for compounds 5 and 6 again suggests that membrane crossing is hydrophobically not important, otherwise π for R_2 would be significant. If DNA is the site of action then it is clear that R_1 can make hydrophobic contact while R_2 does not. It is possible that compounds 1 to 6 penetrate to DNA without dependence on hydrophobicity. With the DNA a hydrophobic interaction could occur by binding between base pairs if the interaction geometry were favorable.

The antitumor and LD_{50} activity of the aniline mustards occurs in rats and mice where there is every opportunity for hydrophobic interactions. Still, in two of the five examples, hydrophobicity plays no role and in the other three there is a small negative dependence on hydrophobicity. If DNA is the site of action in these examples, as is

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

generally believed, one would expect lipophilic compounds to penetrate the nucleus more readily. We believe that it is possible that the selective toxic action in the case of cancer and the toxic action in rats may be occuring with a nucleophile outside of the nucleus in a polar environment.

The good agreement for the ρ values with isolated cholinesterase and toxicity to flies clearly suggests that it is inhibition of the enzyme in the fly which is responsible for the LD_{50}. How the inhibitors reach the enzyme without crossing hydrophobic membranes is not clear. It would seem that either the membrane crossing has a negative effect or that there are hydrophilic channels which provide access to the enzyme.

The apparent radical toxicity of phenols and anilines also does not show positive hydrophobic effects except in the case of the pregnant rat. Such radicals would also be electrophilic agents. In this instance we have no idea of the nature of the species with which the radicals react. It seems unlikely that it is DNA, since the phenols are not active in the Ames test for mutagenicity. Still, it is possible that they could modify DNA without being mutagenic. Of course there are many electron rich molecular moieties with which they could react and it is even possible that more than one could be involved.

We hope that the above survey will attract the attention of others to a problem of importance in the design of bioactive compounds and their toxicity.

Literature Cited

1. Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. *Nature* **1962**, 194, 178-180.
2. Topliss, J.G. *Perspect. Drug Dis. Design* **1993**, 1, 253-258.
3. Hansch, C. *Acc. Chem. Res.* **1993**, 26, 147-153.
4. Hoekman, D.; Leo, A.; Li, P.; Zhang, L.; Hansch, C. In *Proc. 27th Ann. Hawaii Int. Conf. on System Sciences*; Hunter, L., Ed.; IEEE Computer Society Press, 1994, Vol. 5; 193-202.
5. Selassie, C.D.; Hansch, C.; Khwaja, T.A. *J. Med. Chem.* **1990**, 33, 1914-1919.
6. Debnath, A.K.; Shusterman, A.J.; de Compadre, R.R.L.; Hansch, C. *Mutat. Res.* **1994**, 305, 63-72.
7. Hansch, C.; Telzer, B.R.; Zhang, L. *CRC Critical Rev. Toxicol.*, in press.
8. Seydel, J.K. *J. Med. Chem.* **1971**, 14, 724-729.
9. Seydel, J.K. *Mol. Pharmacol.* **1966**, 2, 259-265.
10. Leo, A. *Chem. Rev.* **1993**, 93, 1281-1306.
11. Miller, G.H.; Doukas, P.H.; Seydel, J.K. *J. Med. Chem.* **1972**, 15, 700-706.
12. Hansch, unpublished results.
13. Hansch, C. H.; Deutsch, E. W. *J. Org. Chem.* **1970**, 25, 20-21.
14. Relimpio, A. M. *Gen. Pharmacol.* **1978**, 9, 49-53.
15. Fukuto, T. R.; Metcalf, R. L.; Winton, M. Y. *J. Econ. Entomol.* **1959**, 52, 1121-1127.
16. Metcalf, R. L.; Fukuto, T. R.; Frederickson, M. *J. Agr. Food Chem.* **1964**, 12, 231-236.
17. Fukuto, T. R.; Metcalf, R. L.; Winton, M. Y.; March, R. B. *J. Econ. Entomol.* **1963**, 56, 808-810.
18. Murdock, L. L.; Hopkins, T. L. *J. Agr. Food Chem.* **1968**, 16, 954-958.
19. Fukuto, T. R.; Metcalf, R. L.; Winton, M. Y. *J. Econ. Entomol.* **1961**, 54, 955-962.
20. Sanborn, J. R.; Fukuto, T. R. *J. Agr. Food. Chem.* **1972**, 20, 926-930.
21. Goldblum, A.; Yoshimoto, M.; Hansch, C. H. *J. Agr. Food Chem.* **1981**, 29, 277-288.
22. Darlington, W. A.; Partos, R. D.; Ratts, K. W. *Toxicol. Appl. Pharmacol.* **1971**, 18, 542-547.
23. Fukuto, T. R.; Metcalf, R. L. *J. Am. Chem. Soc.* **1959**, 81, 372-277.
24. Relimpio, A. M. *Gen. Pharmacol.* **1978**, 9, 49-53.

25. Steurbaut, W.; Dejonckreere, W.; Kips, R. H. in *Crop Protection Agents*, McFarlane, Ed. Academic Press, 1977, 79-84.
26. Oglesby, L.A.; Eborn-McCoy, M.T.; Logsdon, T.R.; Copeland, F.; Beyer, P.E.; Kavlock, R.J. *Teratology* **1992**, 45, 11-33.
27. Richard, A.M.; Hongslo, J.K.; Boone, P.F.; Holm, J.A. *Chem. Res. Toxicol.* **1991**, 4, 151-155.
28. Ingold, K.U. *Can. J. Chem.* **1963**, 41, 2816-2825.
29. Mukai, K.; Yokoyama, S.; Fukuda, K.; Uemoto, Y. *Bull. Chem. Soc. Jap.* **1987**, 60, 2163-2167.
30. Harada, A.; Hanazawa, M.; Saito, J.; Hashimoto, K. *Environ. Toxicol. Chem.* **1993**, 11, 973-980.
31. Hansch, C.; Zhang, L. *Drug Metab. Rev.* **1993**, 25, 1-48.
32. Halliwell, B.; Gutteridge, J.M.C. In *Free Radicals in Biology and Medicine*, 2nd Ed., Clarendon Press, 1989, p. 30.
33. Ames, B.N.; Shigenaga, M.K.; Hagen, T.M. *Proc. Natl. Acad. Sci. U.S.* **1993**, 90, 7915-7922.
34. Sussman, J. L.; Harel, M.; Frolow, F; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. *Science* **1991**, 253, 872-879.
35. Selassie, C. D.; Li, R. L.; Poe, M.; Hansch, C. H. *J. Med. Chem.* **1991**, 34, 46-54.
36. Nikaido, H.; Rosenberg, E. J. *J. Bact.* **1983**, 153, 241-252.

RECEIVED May 25, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 15, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch019

Chapter 20

Applications of Scaled Rank–Sum Statistics in Herbicide QSAR

**Robert D. Clark[1,2], John J. Parlow[1], Lawrence H. Brannigan[1],
Dora M. Schnur[1], and David L. Duewer[3]**

[1]Monsanto Company, 800 North Lindbergh Boulevard,
St. Louis, MO 63167
[2]Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144
[3]National Institute of Standards and Technology,
Building 222–Chemistry, Gaithersburg, MD 20899

Quantitative structure/activity relationships (QSARs) can be delineated by multiple regression analysis based on substituent properties, or by comparative molecular field analysis (CoMFA). In either case, the negative logarithms of the doses required to achieve some particular effect (*e.g.*, pED_{50}) have generally been used as response variables. In pesticide development, these must ultimately be obtained from tests on intact organisms. Unfortunately, such parametric point-estimates derived from *in vivo* titration curves may unnecessarily limit the quality of the QSAR obtained. We show here that Scaled Rank Sums (SRS) represent an alternative, non-parametric approach which can provide robust quantitation of pesticidal efficacy suitable for either substituentwise regression analysis or CoMFA.

Any new agrochemical must out-perform chemistry already in the marketplace if it is to be profitable, and must satisfy ever-tightening regulatory and environmental constraints if it is to be registered. The cost of developing new products has risen rapidly in recent years as a result, which has promped pesticide development groups to rely increasingly on delineation of quantitative structure/activity relationships (QSARs) (e.g., (*1-4*)) to increase the pace and efficiency of development. The rapidly falling price of computing power has combined with recent advances in regression techniques to make a wide range of methodologies available to the modern analyst, from classical Hansch substituent-wise regression (*5*) to the more recently introduced comparative molecular field analysis (CoMFA), which correlates biological activity with calculated steric and electronic fields across a molecular lattice (*6,7*). Unfortunately, the usefulness of the regression techniques available is limited by the qualitative properties of the response variables currently available to the researcher. *In vitro* assays involving receptor binding or

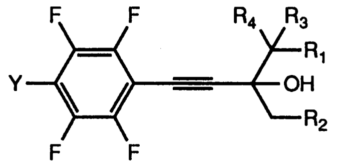
0097–6156/95/0606–0264$12.00/0

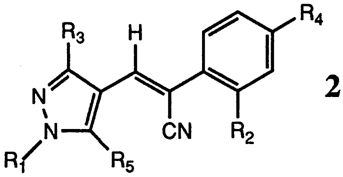
enzyme inhibition and expensive, highly replicated greenhouse studies on single species are well suited to such analyses. But the questions more often faced in an industrial setting involve *in vivo* evaluations of potency across multiple pesticide species. For example, the best commercial candidate among a series of analogs cannot be determined by titration of herbicidal efficacy on any single species.

In rank transformation (RT), parametric response variables such as ED_{50} titration midpoints are replaced by their corresponding ranks. This is one way to mitigate effects of departures from least-squares assumptions and increase the robustness of regression (*8*). Unfortunately, a great deal of resolution is generally lost in applying the rank transform. Moreover, simple underlying algebraic relationships between responses and explanatory (nominally independent) variables, which greatly augment the power of parametric regression, are generally lost under RT. QSARs can be re-linearized by applying RT to the explanatory variables as well (*8*), but this further reduces resolution. In addition, explanatory variables are often only varied across a small set of fixed values - application rates of 0.1, 0.3, 1, 3 and 10 kg/ha being examples from herbicide analysis - so that rates for different treatments are "tied". Such ties reduce the statistical power and robustness of rank analyses.

We have recently introduced Scaled Rank Sums (SRS) as an extension of the rank transform (*9*). In our original report, we noted that Scaled Rank Sums were able to reproduce and extend the scope of expert judgements in greenhouse screens for herbicide interactions. Here, we show that SRS are also useful in delineating herbicide QSARs; applications to data for phenylpropargniols **1** and pyrazole olefinic nitriles **2** are described as particular examples.

1





2

This paper utilizes these two chemistries to explore biometric questions. Only the effects of different manipulations of the raw data on subsequent QSAR results are addressed here. Chemometric interpretation of the results obtained with respect to synthetic follow-up and mechanistic interpretation are considered elsewhere (*10,11*).

Calculations and Simulated titrations

Figures 1A-D show a set of titration curves for hypothetical compounds **I-IV** applied to four species *A-D*. Growth reduction (GR) scores were calculated for ideal logistic titration curves at five application rates between 0.1 and 10 kg/ha and rounded to the nearest 5%. Then a uniformly distributed noise level of ±5% was factored in.

The titrations shown in Figure 1 illustrate some of the fundamental problems faced by a researcher trying to apply QSAR to greenhouse data for evaluating herbicidal potency. The titration range used was not wide enough to allow direct estimation of GR_{50} for compound **I** on species *A*, nor for **II**, **III** or **IV** on species *D*. Indeed, all herbicide titrations bracket the midpoint only for species *B*. But the commercial potential of a herbicide necessarily encompasses activity across a broad spectrum of species, so simply disregarding some species is unacceptable. Note that **II** is appreciably more potent than is **III** only for species *B*, so that an analysis of the *B* data alone would would yield a precise but erroneous answer to the more pertinent question of broad-spectrum activity.

One way to address this dilemma is to average GRs across species for each treatment, as is illustrated in Table I for our hypothetical data set. The resulting titration curves are then far from ideal, however, so that curve fitting becomes impractical. Instead, GR_{50}(Av) and GR_{80}(Av) values cited here were obtained using a χ^2 routine developed at Monsanto (*10*). The value of GR_{50}(Av) obtained is then determined mostly by data for those weeds "median" in sensitivity, whereas GR_{80}(Av) is dominated by activity on those species most difficult to kill, which are generally those most commercially relevant [GR=80% is the minimum commercially acceptable level for control of weed species; GR=20% is the maximum commercially acceptable level of injury to a crop species]. When modest extrapolations must be made, a simple logistic titration curve is assumed.

Rank Sums provide an alternative, non-parametric summary statistic which also draws on data from all species in a test. As the first step in calculating scaled rank sums, RT is applied to the individual GR scores for each treatment in a test. Table I illustrates such a rank transformation for the titration data in Figure 1. Ranks were assigned primarily based on injury to the most sensitive species *A* , but there were eight treatments which completely suppressed growth of *A* (GR(*A*)=100) and three which gave GR(*A*)=90%. These ties were broken by reference to injury on the less sensitive species *B*. There were still ties at this level, however, which were broken by reference to species *C* ; the last remaining tie for GR(*A*)=100 (between ranks 20 and 19) was broken based on the response of species *D* to each herbicide.

The transformed titration curves are plotted as a function of application rate in Figure 2A. The linearization seen here becomes more pronounced as more treatments are included in the ranking, particularly when replication is involved. Although this example involves "pure" logistic curves, ranking at geometric sampling intervals will transform any monotonically increasing (or decreasing) function to a log-linear one. Other functions are linearized around break-points where the first derivative passes through 0. A parabolic function, for example, is transformed to a bi-linear one. This very useful effect of SRS statistics has been noted elsewhere (*9*).

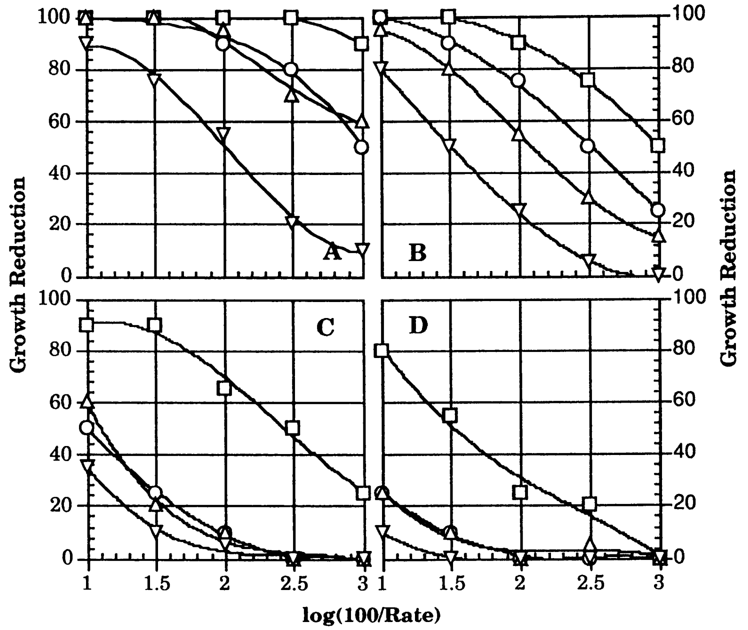


Figure 1: Simulated titrations for four herbicides **I** (□), **II** (○), **III** (Δ) and **IV**(∇) for weed species *A-D*.

TABLE I: Ranks and average GR values for titrations in Figure 1A-D.

	Growth Reduction						Rate	log
Rank[a]	A[b]	B	C	D	Average	Herbicide	(kg/ha)	(Rate/100)
20	100	100	90	80	93	I	10	1
19	100	100	90	55	86	I	3	1.5
18	100	100	50	25	69	II	10	1
17	100	95	60	25	70	III	10	1
16	100	90	65	25	70	I	1	2
15	100	90	25	10	56	II	3	1.5
14	100	80	20	10	53	III	3	1.5
13	100	75	50	20	61	I	0.3	2.5
12	95	55	10	0	40	III	1	2
11	90	80	35	10	54	IV	10	1
10	90	75	10	0	44	II	1	2
9	90	50	25	0	41	I	0.1	3
8	80	50	0	0	33	II	0.3	2.5
7	75	50	10	0	34	IV	3	1.5
6	70	30	0	5	26	III	0.3	2.5
5	60	15	0	0	19	III	0.1	3
4	55	25	5	0	21	IV	1	2
3	50	25	0	0	19	II	0.1	3
2	20	5	0	0	6	IV	0.3	2.5
1	10	0	0	0	3	IV	0.1	3

[a] Rank based on injury to species A, with ties broken by reference to injury on B, C, and D, in descending order.

[b] A, B, C and D correspond to different plant species.

To calculate Rank Sums, treatment ranks are added up *across rates* for each compound tested, and the totals scaled so that the resulting statistic is independent of test size and treatment replication. More formally, the scaled rank-sum SRS is defined by Equation 1:

$$\mathrm{SRS} = \left[\frac{100}{qm(N - qm)}\right] \bullet \left[\left(\sum_{i}^{m}\sum_{j}^{q} U_{ij}\right) - \frac{1}{2}\,qm(qm + 1)\right] \qquad (1)$$

where N is the total number of treatments ranked and U_{ij} are the ranks from 1 (least injury evident) to N (most injury) for q replicates at m rates. The possible SRS values obtained range from a minimum of 0 (all lowest ranks $1,2,\ldots qm$ attributable to the same compound) to a maximum of 100 (all highest ranks N, N-1,...N-qm arising from one compound). Formally, this makes scaled-rank sum statistics an extension of the Wilcoxon and Mann-Whitney tests (*12,13*) to allow multiple comparisons among related sets of observations.

Note that N may be larger than nqm, where n is the number of compounds being evaluated. The number of titration "steps" considered in calculating the rank-sum for each compound must be consistent, however. Ratings for check pans treated only with formulation and solvent can also be included; this is particularly useful in fungicide evaluations, where some treatments may in fact encourage disease.

Logistic curves have a common shape on a log-linear scale, differing only in their offsets (midpoints) along the abscissa. As a result, the rank transforms in Figure 2A are essentially parallel. The difference in Scaled Rank Sums (ΔSRS) for two compounds **a** and **b** at m rates is then given by Equation 2, wherein v_i's represent the negative logarithms of the application rates used, β is the magnitude of the common slope for the rank-transformed titration curves, and U^0's correspond to the respective ordinate intercepts.

$$\mathrm{SRS}_a - \mathrm{SRS}_b = \left[\frac{100}{m(N-m)}\right] \bullet \left[\sum_{i}^{m}(U_a^0 - \beta v_{ia}) - \sum_{i}^{m}(U_b^0 - \beta v_{ib})\right] \qquad (2)$$

$$= \left[\frac{100}{m(N-m)}\right] \bullet \left[m\,(U_a^0 - U_b^0) - \sum_{i}^{m}\beta\,(v_{ia} - v_{ib})\right] \qquad (3)$$

Grouping similar terms in Equation 2 together gives Equation 3. But, in this case, $U^0 = \beta v^0$, where v^0's are the v-axis intercepts. If a constant offset Δv_{ab} is maintained between **a** and **b** across the titration range, the above can be re-arranged to yield Equation 4:

$$\Delta\,\mathrm{SRS} = \alpha\,(v_a^0 - v_b^0 - \Delta v_{ab}) \qquad (4)$$

where $\alpha = 100\beta/(N - m)$. Equation 4 is central to the application of scaled rank-sums. It implies that when titrations are all carried out at the same set of rates ($\Delta v = 0$, as in Figure 1), SRS will be linearly related to potency, whether measured at the midpoint (GR_{50}) or at some other fixed threshold such as GR_{80}. Moreover, titrating a single compound (*i.e.*, **a** = **b**) at offset rates (*e.g.*, 0.1, 0.3 and 1 *versus* 1, 3 and 10kg/ha; $\Delta v = \log 10 = 1$) will provide a direct estimate of α, which can then be used to correct SRS values for other compounds assayed at offset application rates. This is illustrated below in the PONS herbicide analysis.

In Figure 2B, scaled rank sums have been plotted for five sets of curves as a function of GR_{50}. Within each set, fifteen GR_{50} values were selected from a log-uniform, random distribution ranging from 0.01 to 100 kg/ha. For each set, GRs at 10, 3, 1, 0.3 and 0.1 kg/ha were calculated and the scaled rank sums calculated as described above. Note that all five sets exhibit slopes near 20 per log unit (*i.e.*, 20/decade); so long as the titration range and weed species used do not limit dynamic range, the slope α falls near $80/Q$, where Q is the common logarithm of the range in potencies examined (here, 4 decades).

In our experience, three species are enough to resolve ties, particularly when they are considered in order of decreasing (or increasing) sensitivity.

Case I: Phenylproparginols

Twenty congeneric phenylproparginols **1** were examined in a side-by-side greenhouse test to obtain both parametric (GR_{80}) and SRS measures of herbicidal activity. Space limitations preclude setting out the individual structures here; interested readers are encouraged to consult the original publication for details (*10*).

The biological assay used for phenylproparginols is also detailed elsewhere (*10*). The weed species tested included barnyardgrass (*Echinochloa crus-galli*; BYGR), yellow nutsedge (*Cyperus esculentus* L.; YENS), yellow foxtail (*Setaria glauca*; YEFT), seedling johnsongrass (*Sorghum halopense*; SEJG), large crabgrass (*Digitaria sanguinalis*; LACG) and downy brome (*Bromus tectorum*; DOBR). Applications were made as duplicate pre-emer-gence treatments at 2.5, 0.5 and 0.1 kg/ha. GR_{80}s fell outside the test's dynamic range for at least one analog for most of the weeds tested, so pGR_{80}s could not be directly averaged across species Instead, $GR_{80}(Av)$ values cited here refer to injury averaged across species at each rate (*10*). Here and throughout this paper, pGR values were referenced to 100 kg/ha, *i.e.,* pGR80 = log(100/GR80), where GR80 is given in kg/ha (*14*).

Figure 2B showed that a log-linear relationship exists between SRS and GR_{50} for ideal titrations; Figure 3A, which plots SRS for phenylproparginols against GR_{80}s, shows that this same relationship holds for authentic greenhouse data. Non-parametric linear regression (*15*) gave an estimate of 42/decade for the crabgrass slope α shown in Figure 3A.

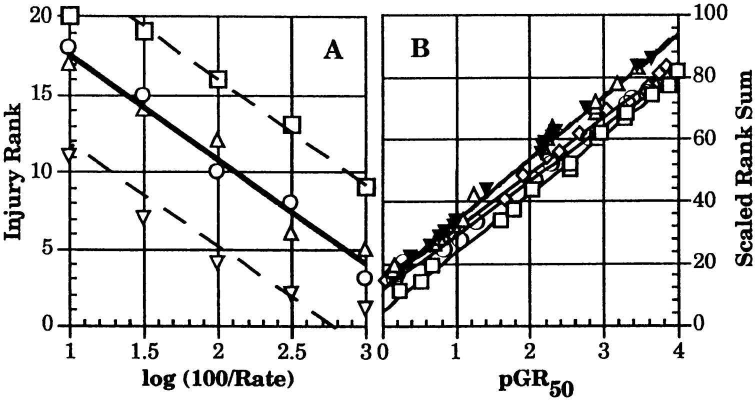


Figure 2: (**A**) Titration curves for data ranked as specified in Table I. (**B**) Correlation between scaled rank-sum statistics and pGR_{50} for five families of titration curves, each calculated for fifteen randomly generated pGR_{50} values lying between 0 and 4. Simulated titrations included five rates between 0.1 and 10 kg/ha. Each line and type of symbol corresponds to one family of titration curves.

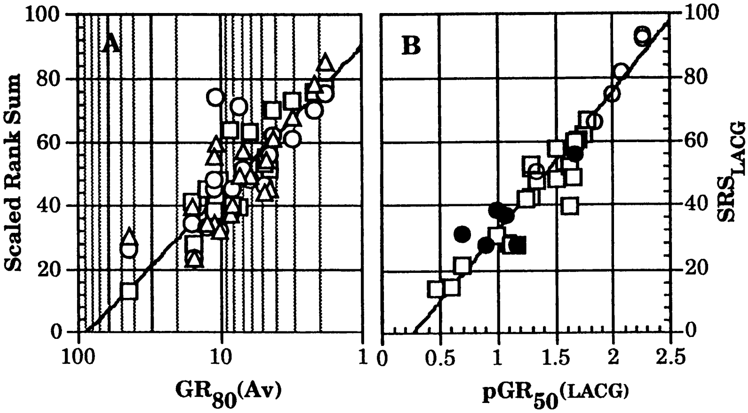


Figure 3: Log-linear correlation between Scaled Rank Sums and parametric measures of potency. (**A**) SRS_{BYGR} (△), SRS_{DOBR} (○) and SRS_{YENS} (□) for phenylpropanginol herbicides as a function of pGR_{80}(AV). Ranking Criteria for ranking for SRS_{YENS} were **YENS** > YEFT > BYGR; other priorities were as given in Table I. The line shown has a slope α of 42/decade. (**B**) SRS_{LACG} as a measure of PONs potency *versus* pGR_{50}(LACG). Open square symbols (□) correspond to titrations limited to between 1 and 10 kg/ha. Circles correspond to results from extended titration series: (○) 1-10 kg/ha; and (●) 0.1-1 kg/ha. The latter are plotted as though for 10X dilutions. The line shown has a slope α of 44/decade.

Departures from exact co-linearity in Figure 3A are due partly to random noise in both response variables. But the differences also reflect underlying systematic differences which could tend to make one or the other statistic more useful as a response variable. The latter possibility was explored by examining QSARs derived with one or the other statistic as dependent variable.

Hansch analysis. The group of phenylproparginols analyzed here includes an experimental design set balanced across three property descriptors for phenyl substituent Y and substituents R_1 and R_2 geminal to the hydroxyl group (R_3 and R_4 =H); a few analogs with tertiary and quaternary α-carbons geminal to the hydroxyl group (R_3 and/or R_4 = Me) were also included (*10*). Syntheses were not carried out enantioselectively, however, and products were tested as racemic mixtures. The stereochemical ambiguities introduced confound the design somewhat. For purposes of analysis, the less substituted α-carbon was designated as R_1 and the multiple regressions were run using only Hansch's lipophilicity parameter π ($\equiv \pi_1$) as descriptor for this position; electronic (σ^*) or steric (E_S) substituent parameters (*16,17*) would have been algebraically equivalent to π, though the magnitudes of the coefficients obtained would have differed. Steric, electronic and hydrophobicity parameters were all included for substituents R_2 and Y, however. Inicator variables I_3 and/or I_4 were set to 0 or 1 to indicate whether R_3 or R_4 was H or CH_3, respectively.

Squared terms were tested for significance after purely linear models had been optimized. Only those quadratic terms which were significant *and* which corresponded to optima were incorporated into the models. Such terms were centered about their respective optima in x_j (Equation 5) and the analyses reiterated.

$$\delta y/\delta x_j = a_j - 2b\hat{x}_j = 0 \Rightarrow \hat{x}_j = -a_j/2b_j \qquad (5)$$

Table II lays out the results for such ordinary least-squares (OLS) analysis of the phenyproparginol data.

Only SRS_{BYGR} and pGR_{80}(LACG) gave stable models when the response variables were evaluated independently. Of these two, SRS_{BYGR} gave the better model. This was so whether one used Pearson's r^2 as a criterion or one looked to the standard error of regression (SE); in the latter case, the 42/decade proportionality factor α from Figure 2A had to be taken into consideration. In addition, the SRS model was more descriptive, in that it characterized contributions from $E_S(2)$ and σ_Y, terms which did not contribute significantly to the pGR_{80} model.

The parameter estimates obtained differ somewhat between the local SRS_{BYGR} and pGR_{80}(LACG) models set out in Table II. The differences in estimated coefficients are marginal both statistically and with respect to setting synthetic direction. The optima found for π_1 are reassuringly similar, but it is important to know the "true" value of the optimum for $E_S(Y)$ and whether the term in σ_Y does or does not affect activity.

The SRS model was found to be the more stable of the two when the respective deduced $E_S(Y)$ optima were interchanged. This interchange produced the mixed models given in Table II. The fit improved slightly in terms of r^2 for the pED_{80} model, whereas both the residual error (SE) and the r^2 for the SRS model suffered appreciably. The latter lost its term in σ_Y, whereas the pED_{80} model gained terms in both σ_Y and σ_2. These results confirm the more robust nature of the model based on SRS.

Case II: Pyrazole Olefinic Nitriles (PONs)

Twenty analogs of a lead pyrazole olefinic nitrile were designed so as to be variously substituted at the 2' and 4'-positions of the phenyl ring and at the 1, 3 and 5 positions of the pyrazole ring, in an experimental design similar to that outlined previously for chloroacetanilides (*14*). The structures and summary herbicidal potency data for fourteen of these PONs (**2a-n**) are set out in Table III; more complete data are available from the authors. Those target structures in the original design not included here bore acyl substitutions at N^1, the lability of which *in vivo* would have complicated interpretation of the results obtained.

The biological activities of PONs herbicides were assayed much as were phenylproparginols, except that annual bluegrass (*Poa annua*; ANBG) was included in the PONs tests, whereas yellow nutsedge, yellow foxtail and downy brome were not. In addition, PONs herbicides were tested at the somewhat higher application rates of 10, 3.2, and 1.0 kg/ha; 0.32 and 0.10 kg/ha treatments were added for more active analogs **2e**, **2f** and **2k**. Smaller pots were used than for phenylproparginols (10x10 cm instead of 11x27 cm).

Crabgrass GR_{50}s could be estimated accurately for all PONs analogs tested. Other species exhibited growth reductions of 10% or less for one or more analogs, so GR_{50}s could not be extrapolated for those compounds on those individual species with any confidence. Hence only the crabgrass (LACG) and average (Av) pGR_{80}s are included in Table III. The tiered system of ranking criteria discussed above allows the accurate calculation of rank sums even when some compounds are inactive on some species, however, so SRS statistics were determined with either crabgrass or johnsongrass (SEJG) as primary sorting criteria. As can be seen in Table III, both rankings gave broadly similar results.

As is illustrated in Figure 3B, the log-linear relationship between GR_{50} and SRS on crabgrass holds even more strictly for PONs than for phenylproparginol herbicides. The slope α was estimated by non-parametric linear regression to be 44/decade. Data are also shown for titrations in which the rate was extended by 1 log unit for **2e**, **2f** and **2k** (● in Figure 3B *vs* O). This corresponds to $\Delta v = 1$ in Equation 4; the relation defined

TABLE II: Ordinary least squares regression analysis for herbicidal phenylpropargninols **1a-t**.

$$y/\alpha = \{1 + \Sigma a_i x_i - \Sigma b_j(x_j - \hat{x}_j)^2\} \bullet c \qquad \text{[Equation 2]}^a$$

	SRS$_{BYGR}$						log[100/GR$_{80}$(LACG)]					
	local			mixed			local			mixed		
x_i	a_i	b_j	$\hat{x}_j$	a_i	b_j	$\hat{x}_j$	a_i	b_j	$\hat{x}_j$	a_i	b_j	$\hat{x}_j$
I_3	0.389 (±.093)[b]	---	---	0.193 (±.092)	---	---	0.221 (±.098)	---	---	0.440 (±.143)	---	---
π_1	---	1.022 (±.200)	0.65	---	0.760 (±.221)	0.65	---	0.848 (±.313)	0.60	---	1.196 (±.404)	0.60
$E_S(Y)$	---	1.456 (±.384)	0.93	---	0.350 (±.157)	**0.57**[c]	---	0.676 (±.154)	0.57	---	1.441 (±.578)	**0.93**
$E_2(Y)$	–0.158 (±.054)	---	---	–0.138 (±.060)	---	---	---	---	---	---	---	---
σ_Y	0.240 (±.075)	---	---	---	---	---	---	---	---	0.288 (±.114)	---	---
σ_2	---	---	---	---	---	---	---	---	---	–0.132 (±.057)	---	---
c		1.78			1.93			2.11			1.71	
SE[d]		±0.17			±0.21			±0.25			±0.26	
r^2		0.835			0.740			0.756			0.779	

[a] Scaling factor α = 1 for GR$_{80}$ data; α = 42 for scaled rank-sums (SRS). [b] Parentheses set off standard errors for coefficients. [c] Boldface indicates extraneously supplied parameters. [d] Root mean square deviation from regression.

TABLE III: Structures and biological response values for PONs herbicides **2a-n**.

						Scaled Rank Sums		log[100/GR$_{50}$][a]	
Cmpd	R_1	R_2	R_3	R_4	R_5	LACG	SEJG	LACG	Average
2a	Me	F	H	H	CF_3	17	36	0.59	0.76
2b	Me	F	iPr	F	SMe	59	64	1.61	1.22
2c	Et	H	Me	CF_3	Cl	17	18	0.80	0.00
2d	iPr	CF_3	Cl	H	Me	52	54	1.56	1.31
2e	Et	CF_3	CF_3	H	H	93	94	2.28	2.15
2e/10[b]						[39]	---	[1.43][b]	[1.13]
2f	Me	F	CF_3	H	CH_2CN	56	57	1.59	1.23
2f/10[b]						[42]	---	[0.80]	[0.30]
2g	Et	F	H	H	Me	34	28	1.37	0.86
2h	iPr	H	iPr	F	Cl	38	28	1.38	0.67
2i	iPr	Cl	Me	H	H	55	53	1.63	1.32
2j	Et	F	Me	F	CF_3	47	40	1.27	1.11
2k	iPr	CF_3	Me	H	H	78	84	2.04	1.77
2k/10[b]						[37]	---	[1.04]	[0.76]
2l	Me	CF_3	iPr	H	CF_3	39	46	1.24	1.06
2m	iPr	F	Cl	H	CF_3	56	59	1.66	1.40
2n	Me	Cl	CF_3	H	CH_2CN	45	39	1.46	0.88
SEM						±6.82	±0.16	±0.12	±6.41
SEM/α[c]						±0.16	±0.15	±0.16	±0.12
SEM/range						±.090	±.084	±.095	±.056
F^{13}_{14}						8.66	10.22	7.26	17.63

[a] Based on data from 10, 3 and 1 kg/ha treatments except as noted. GR$_{50}$s above 2 and below 1 were estimated by extrapolation assuming logistic titrations. [b] Based on data from 1, 0.3 and 0.1 kg/ha treatments, *i.e.*, for 10x dilutions. [c] Scaling factor α = 1 for GR$_{50}$ data, α = 44 for SRS statistics.

by these sets of points is gratifying close to that defined by SRS *vs* pGR_{50}. Such internal calibration is generally more useful than are SRS/pGR_{50} plots for determining correction factors α to compensate for the effects of differences in application rates on Scaled Rank Sums.

Precision. Standard errors (SEM) were estimated by analysis of variance across replicate titrations. The error in average weed pGR_{50} was somewhat smaller than for pGR_{50} for crabgrass alone, whereas SRS_{SEJG} had a smaller sampling error than did SRS_{LACG}. Errors for the two classes of response variable (pGR and SRS) cannot be compared directly, because they fall on very different measurement scales. Relative precisions can be compared, however, by adjusting for the individual scales involved. For these data, one reaches the same conclusion whether error is considered relative to the variation in means (F statistic); relative to the range of means; or after scaling by the slope α estimated from Figure 4B. In any case, both Scaled Rank-Sums were slightly more precise than was pGR_{50}(LACG), whereas pGR_{50}(Av) was more precise than either.

Although precision may limit a response variable's usefulness in QSAR, the most precise response variable is not necessarily the best suited to regression analysis. Even a very precise response variable may depend on descriptors in a way that seriously compromises QSARs based on it. Measurement error can propagate very differently across different classes of response statistics, so that precision cannot in and of itself be an adequate basis for comparison. Our experience indicates that pGR_{80}s and pGR_{50}s are at best imperfect as response variables for least-squares regression in general and for characterizing free-energy response surfaces in particular. Scaled-Rank Sums are usually better suited to such analysis.

The PONs data provide a case in point. When subjected to a Hansch analysis similar to that described above, pGR_{50}(LACG) returns the most precise models of the four response statistics examined. As for phenylproparginols, however, both SRS statistics give "richer" models than are obtained for either pGR_{50} (details not shown).

CoMFA. The flexibilty of phenylproparginols and the chiral center they bear complicates any application of comparative molecular field analysis (CoMFA), a complement to Hansch analysis which allows greater variation in molecular structure and which can take into account interactions in three dimensions (*6*). The PONs herbicides, on the other hand, have a relatively rigid framework and small substituents. This makes them good candidates for CoMFA because it minimizes effects of small diffrences in alignment between molecules.

Three dimensional molecular structures were generated in SYBYL 5.47 from SMILES strings using CONCORD 2.8. Comparable conformations for all analogs were generated by manual bond rotations, the local conformational energy minimum in each instance being calculated using an AM1 Hamiltonian in MOPAC 5.0. The "relaxed" structures obtained

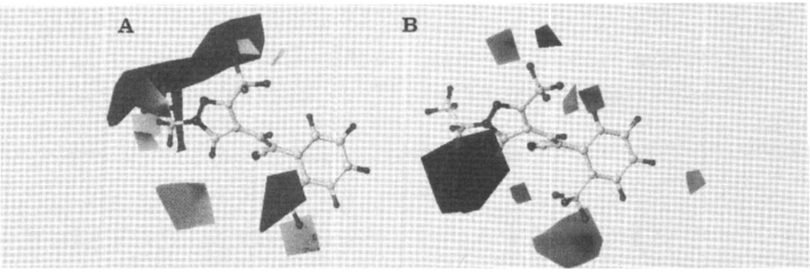


Figure 4: CoMFA maps for PONs analogs (pGR$_{50}$(LACG)). The structure of compound **2e** is included for reference. (**A**) Steric map. Darker shading indicates regions where adding steric bulk enhances herbicidal activity, and lighter shading corresponds to regions in which adding steric bulk is reduces herbicidal activity. CoMFA contour thresholds were 65 and 30%, respectively. (**B**) Electronic map. Darker shading indicates regions in which adding electronegative substituents decreases herbicidal activity, and lighter shading indicates regions in which electronegative substituents tend to give higher activity. CoMFA thresholds were 80 and 25%, respectively.

were aligned with one another by least-squares fitting of atoms in and around the olefinic bond. The steric and electronic fields required for CoMFA were then calculated as the respective energies for an sp^3-carbon atom bearing a unit positive charge situated on interstices of an eveloping lattice. Such energies were calculated at 2 Å intervals along the lattice. Correlations of the resulting fields with herbicidal activity were then extracted by partial least squares (PLS) analysis.

CoMFA maps obtained for SRS$_{\text{LACG}}$ are shown in Figure 4 to illustrate the general results obtained. Large substituents are favored (darker shading in Figure 4A) at R$_1$ and at R$_3$, whereas small substituents are favored for R$_5$ (lighter shading). The presence of electronegative groups at R$_2$ is correlated with high herbicidal activity (lighter shading in Figure 4B), whereas relatively electropositive groups are preferred (darker shading) at R$_1$ and R$_5$. The thrust of the present work, however, is consideration of the suitablity of different types of statistics as response variables, not discussion of specific features of the particular QSAR models obtained.

Precision in CoMFA. The actual descriptors in CoMFA are orthogonal linear combinations (components) of field intensities estimated at various points around the molecules in question. Unfortunately, this makes comparing CoMFA models less straightforward than comparing Hansch equations. The structural orthogonality built into experimental designs of the sort used here further complicates matters. Hence the suitability of parametric and rank-sum statistics as response variables in CoMFA were compared by creating separate CoMFA models from each of the duplicate titrations, randomly assigning duplicates for each treatment to one or the other data set. Correlations r_x^2 were then determined between the corresponding lattice coefficients in the two models.

Results for two-, four- and six-component models are given in Table IV in terms of r_x^2 for steric and electronic coefficients across duplicate models. The values obtained are comparable for all four response variables, with SRS_{LACG} performing best overall. Note, too, that only for SRS_{LACG} was the more complete six-component model also the most precise of those examined.

The standard errors of regression are also given in Table IV. Except for the most reduced, two-component models, all are over-specified (compare the SEs with those given in Table III). A better measure of model precision for PLS is how well response data from one set of titrations can be predicted from the responses observed in the duplicate titration set. This was calculated as the root mean square deviation from prediction between duplicate models, denoted SE_x in Table IV; q^2 (cross-validated r^2) for this method of cross-validation is also given. Five to six PLS components had to be brought into the analysis before all compounds were represented; this is to be expected for a set of compounds designed so as to be maximally diverse in structure, and agrees with the underlying degrees of freedom identifiable by Hansch analysis (data not shown).

By this criterion, the reduced models are most precise across all response variables, with pGR_{50}(LACG) giving the most precise model. But the reduced models are weak, in that they effectively exclude any contribution from several compounds. At the more broadly-based, six-component level, CoMFA precision parallels the precision of the underlying data (Table II), *i.e.*, $pGR_{50}(Av) \geq SRS_{SEJG} \geq SRS_{LACG} \geq pGR_{50}(LACG)$

Another criterion for evaluating the suitability of various response variables for QSAR is how well model assumptions are met. The cross-validated deviations of the SRS statistics examined here are more normally distributed than are the cross-validated deviations based on pGR_{50}s. That is, the skewness and kurtosis (g_1 and g_2, respectively) of the pGR_{50} cross-residuals are generally greater than are the corresponding values for the SRS statistics (Table IV).

The sharpest distinction among the different models, however, is qualitative. Figures 5 and 6 show the pairs of CoMFA fields obtained from replicate experiments. These have been rotated slightly from the views in Figure 4 to make the duplicate maps easier to compare, and the molecular structure has been omitted to de-emphasize the features of the models themselves. Both the steric and electrostatic maps are more similar for the replicate SRS-based model than are either pGR_{50}-based model pair. The replicate pGR_{50}(Av) electrostatic maps shown in Figure 6C, for example, differ substantially. The corresponding steric maps in Figure 5C are also diverge substantially from one another, whereas the SRS maps in Figure 5B are more similar.

TABLE IV: Summary of CoMFA results for PONS herbicides.

	Scaled Rank Sums						**log[100/GR$_{50}$]**					
	LACG			**SEJG**			**LACG**			**Average**		
components	2	4	6	2	4	6	2	4	6	2	4	6
r_x^2 ster[a]	0.87	0.67	0.92	0.83	0.75	0.75	0.77	0.48	0.81	0.84	0.77	0.71
r_x^2 ele[b]	0.82	0.83	0.95	0.89	0.91	0.77	0.89	0.79	0.94	0.92	0.94	0.89
SE[c]	±0.18	±0.08	±0.03	±0.20	±0.09	±0.04	±0.17	±0.08	±0.03	±0.17	±0.08	±0.04
SE_x[d]	±0.15	±0.23	±0.21	±0.13	±0.21	±0.19	±0.14	±0.26	±0.22	±0.15	±0.16	±0.17
q^2	0.89	0.75	0.78	0.92	0.80	0.83	0.89	0.62	0.71	0.90	0.90	0.88
$res_x\ g_1$[e]	-.486	-.097	-.187	-.953	.048	-.047	-.448	-.687	-.402	-1.12	-.697	-.988
$res_x\ g_2$[f]	.159	.153	-.692	1.35	-.903	-.952	-1.15	.098	-.722	1.69	.244	.924

[a] Correlation of steric coefficients between replicate models. [b] Correlation of electronic coefficients between replicate models. [c] Root mean square deviation from regression within models. [d] Cross-validated RMS deviation between models. [e] Skewness of deviations from cross-validated predictions. [f] Kurtosis of deviations from cross-validated predictions.

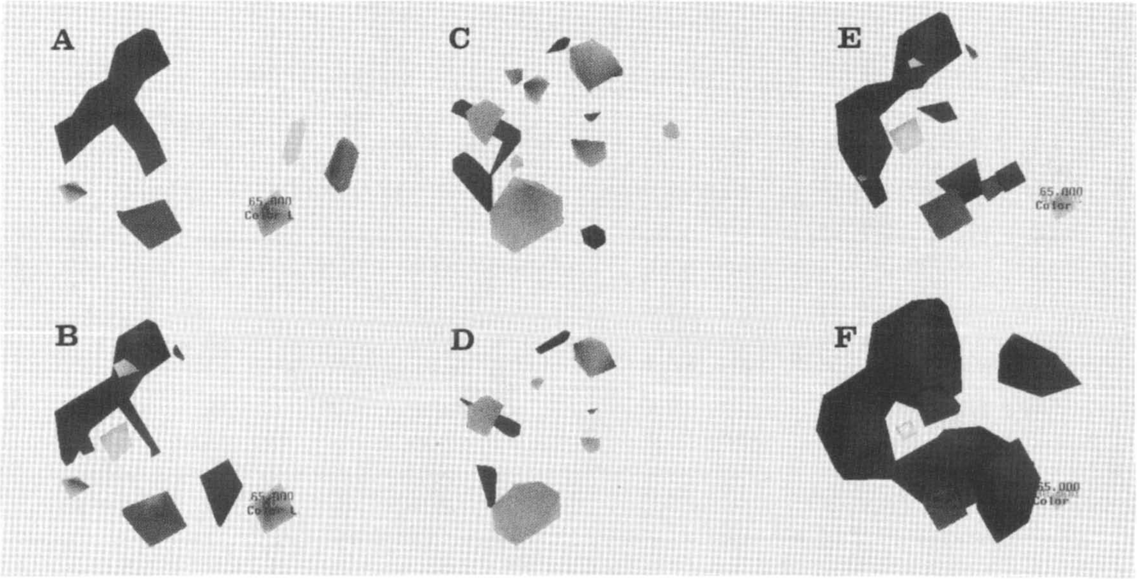

A
C
E
B
D
F
65.000
Color
65.000
Color
65.000
Color
65.000
Color

Figure 5: Comparison of replicate steric CoMFA maps for PONs analogs. (**A**,**B**) Steric maps for pGR_{50}(LACG); (**C**,**D**) for SRS_{LACG}; and (**E**,**F**) for pGR_{50}(AV). Contour thresholds are as indicated in Figure 4.

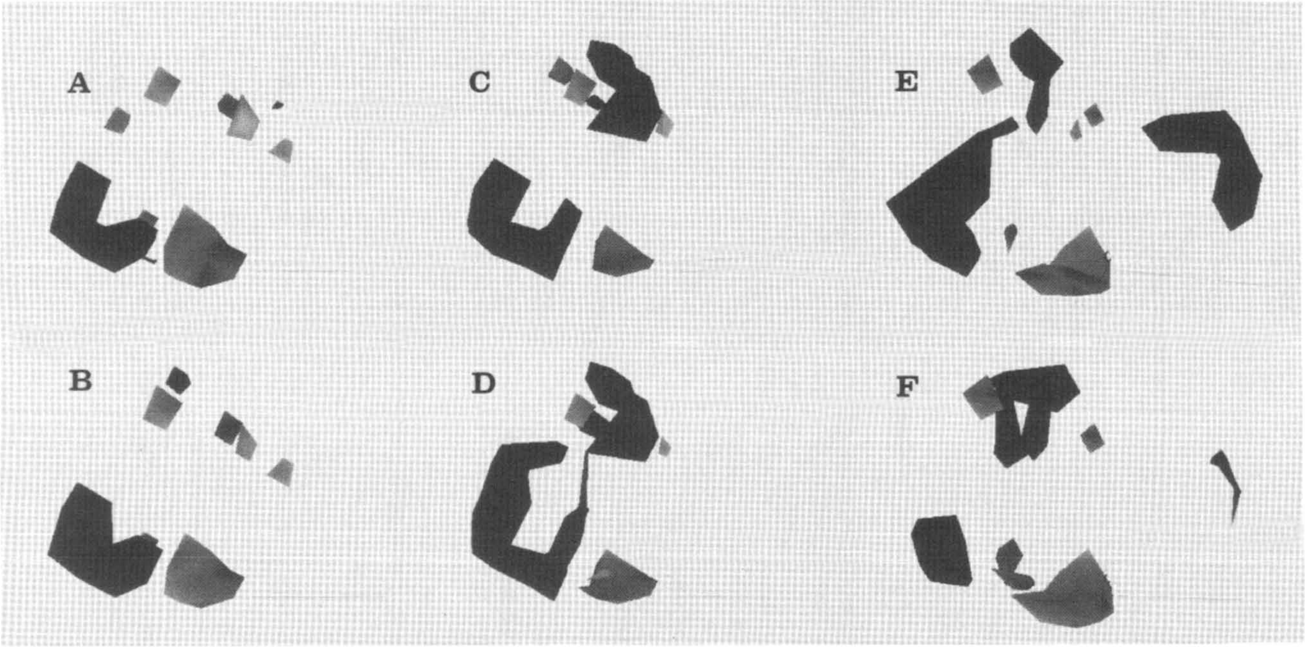


Figure 6: Comparison of replicate electrostatic CoMFA maps for PONs herbicides. Layout is as described for Figure 4.

Discussion

In classical herbicide QSAR, conversion from GR_{50} (or GR_{80}) to pGR_{50} (pGR_{80}) serves to attenuate differences in variance between potent (low GR_{50}) and weakly active (high GR_{50}) compounds, so that the common variance (homoscedasticity) assumptions of least-squares regression can be better satisfied. But experience has shown that, in addition, pGR_{50}s often behave as though they reflect a free energy of herbicide interaction with an intact organism. That is to say, *in vivo* structure/activity relationships can often be described to a good approximation by linear or parabolic functions of substituent parameters derived from simple chemical reactions. Rank transformed GR_{50}s (or GR_{80}s) lack this linear free energy character, and this lack overwhelms clear advantages the rank transform has with respect to robustness and precision. Here we have shown that Scaled Rank Sums capture the linear free-energy flavor of pED_{50}s without sacrificing precision.

Scaled Rank Sums as implemented here differs fundamentally from both GR_{50}s and rank-transformation of parameters in allowing consolidation of data across several species into a single response variable. The algorithm applied entails a natural non-linear weighting across species such that each species contributes to the SRS in direct relation to its information content. For example, the SRS for the most potent compounds is determined primarily by the activity on the least sensitive species, whereas resolution among the least potent compounds is determined by activity on the most sensitive species. Hence using Scaled Rank Sums broadens the effective dynamic range of most titration series.

The log-linear relationship between Scaled Rank Sums and GR_{50} (and GR_{80}) makes it simple to convert the former to more familiar "kg/ha" units. It also makes practical direct use of the more robust SRS values in regression analysis. Even for classically well-behaved cases such as those described here, SRS models behave as well or better than those obtained from pGR_{50}s and pGR_{80}s. In more typical, less ideal cases, the SRS models consistently out-perform their parametric counterparts.

Changes at any of the varied positions along the phenylproparginol and PONs skeletons affect herbicidal activity to some extent, so the more well-determined coefficients any model identifies, the better it can reflect reality. There is a trade-off involved: the more complex a model is, the greater the opportunity that extraneous descriptors will be included. But identifying more descriptors can be very useful when laying out designs for synthetic follow-up even when they are subsequently shown to be extraneous. It is much easier, after all, to test any hypothesis, including an erroneous one, than it is to test an hypothesis which has not been formulated at all.

Acknowledgements

The authors would like to thank the editors and our anonymous reviewers for their careful consideration, insightful comments and constructive suggestions; this article benefited substantially from their input.

References

1. Lyga, J.W.; Patera, R.M.; Theodoridis, G.; Halling, B.P.; Hotzman, F.W.; Plummer, M.J. *J. Agric. Food Chem.* **1991**, *39*, 1667-1673.
2. Nandihalli, U.B.; Duke, M.V.; Duke, S.O. *J. Agric. Food Chem.* **1992**, *40*, 1993-2000.
3. Osabe, H.; Morishima, Y.; Goto, Y.; Masamoto, K.; Nakagawa, Y.; Fujita, T. *Pestic. Sci.* **1992a**,*34*, 27-36.
4. Simmons, K.A.; Dixson, J.A.; Halling, B.P.; Plummer, E.L.; Plummer, M.J.; Tymonko, J.M.; Schmidt, R.J.; Wyle, M.J.; Webster, C.A.; Baver, W.A.; Witkowski, D.A.; Peters, G.R.; Gavelle, W.D. *J. Agric. Food Chem.* **1992**, *40*, 297-309.
5. Hansch, C.; Fujita, T.A. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.
6. Cramer, R.D., III; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.*, **1988**,*110*, 5959-5967.
7. Osabe,H.; Morishima, Y.; Goto, Y.; Fujita, T. *Pestic. Sci.* **1992b**, *35*, 187-200.
8. Iman, R.L.; Conover, W.J. *Technometrics* **1979**,*21*, 499-509.
9. Duewer, D.L.; Clark, R.D. *J. Chemometrics* **1993**, *5*, 503-521.
10. Parlow, J.J.; Clark, R.D. *J. Agric. Food Chem.* **1994**, *42*, 2600-2609.
11. Brannigan, L.H. Monsanto Company, unpublished results.
12. Wilcoxon, F. *J. Biometrics* **1945**, *1*, 80-83.
13. Mann, H.B.; Whitney, D.R. *Ann. Math. Statist.* **1947**, *18*, 50-60.
14. Brannigan, L.H.; Duewer, D.L. *Pharmacochem. Libr.* **1991**, *16*, 553-556.
15. Brown, G.W.; Mood, A.M. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* ; Neyman, J., Ed.;. University of California Press: Berkeley and Los Angeles, 1951.
16. Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: New York, 1978.
17. Hansch, C.; Leo, A.; Taft, R.A. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.*, **1991**, *91*, 165-195.

RECEIVED May 25, 1995

Chapter 21

Novel Structure–Activity Insights from Neural Network Models

Tariq A. Andrea

Stine-Haskell Research Center, DuPont Agricultural Products
P.O. Box 30, Newark, DE 19714

In regression QSAR, ligand/protein binding is a linear or parabolic function of ligands' physico-chemical properties. Due to the absence of higher order and cross-product terms, dependence of binding on one property is invariant to others. By comparison, neural networks are capable of delineating highly non-linear features. Stepwise regression of benzene sulfonylureas' binding to acetolactate synthases enzyme indicates that, properties of the ortho substituent R_2 have the following order of significance: MR > π > $\mathcal{F}$. Affinity depends parabolically on MR (MR_{opt}=14.05) and increases linearly with π and $\mathcal{F}$. The constant curvature (-0.0158) of the MR parabola indicates that the binding pocket tolerates R_2's with 4-7 heavy atoms and that this tolerance is invariant to π and $\mathcal{F}$. This suggests an enzyme pocket with fixed size. Neural networks analysis finds the same order of significance of R_2 properties. While in this model MR dependence is not mathematically parabolic, it has a "parabolic or Guassian-like" shape. Like regression QSAR, the neural model indicates that optimal MR and tolerance to size variation depends on π and $\mathcal{F}$. It suggests a binding pocket which accommodates larger hydrophobic and electron withdrawing substituents than hydrophilic and electron donating ones. It also indicates higher tolerance to size variations in hydrophobic and electron withdrawing substituents than in hydrophilic and electron donating ones. These are consistent with x-ray crystallographic findings that even structurally related ligands can bind differently to the same protein.

In the original formulation of the field of quantitative structure-activity relationships (QSAR), introduced in the early 1960s by Hansch and co-workers, biological activities of chemical compounds are linear or parabolic polynomial functions of their physico-chemical properties (π, MR, σ ...).

$$\text{Activity} = a + b\,\pi + c\,\pi^2 + d\,\text{MR} + e\,\text{MR}^2 + f\,\sigma + \ldots + \ldots \qquad (1)$$

0097–6156/95/0606–0282$12.00/0

Higher order non-linear and cross products terms are not used in practice. Multiple linear regression is used to calculate the a, b, ... coefficients . Back Propagation neural networks were introduced in QSAR a few years ago (*1,2,3*). The initial thrust was to compare their predictions with those of regression QSAR. These publications alluded to, albeit did not expound, the non-linear modeling capabilities of neural networks.

The objective of this publication is to demonstrate the ability of neural networks to delineate non-linear relationships between biological activity and physico-chemical properties and to explicate the novel insights they provide. This theme is developed using data from the inhibition of acetolactate synthase (ALS) enzyme by herbicidal sulfonylureas(*4*).

Data and Model Development

Table 1 (*5*) shows structures and ALS enzyme inhibitory activities of benzenesulfonylureas used in calculating regression and neural QSAR's. I_{50} is the molar concentration for 50% enzyme inhibition. Log ($1/I_{50}$) was modeled as a function of π (Hansch's hydrophobicity parameter), MR (molar refractivity), and Swain-Lupton's electronic parameters $\mathcal{F}$ and $\mathcal{R}$ (*6*) of R_2. Numerically positive values of $\mathcal{F}$ (inductive) and $\mathcal{R}$(resonance) represent electron withdrawing effects and vice versa. The experimentally determined standard deviation of log ($1/I_{50}$) of ~ 0.3 is the infemum (least upper bound) for models' variances. All models with variance ≤ 0.3 are equally valid.

As usual, classical QSAR's were calculated using stepwise multiple linear regression to fit ALS enzyme inhibitory activities to physico-chemical properties. This served to elucidate the relative significance of each independent variable in explaining enzyme inhibitory activity as determined by regression analysis.

A standard three layer back propagation neural network (Figure 1) was used for calculating neural models using π, MR, $\mathcal{F}$ and $\mathcal{R}$as inputs and enzyme inhibition (log($1/I_{50}$)) for the output. These were trained with the stiff differential equations algorithm of Owens and Filkin (*7*). Following the protocols of reference 2, the dataset in Table 1 was randomly split into training and test sets. The test set consisted of 9 data points (entries 4, 8, 12, 16, 20, 24, 28, 32, and 36 of Table 1) while the remaining 30 entries served as the training set. Each neural model was trained to fit ALS enzyme inhibitory activities of the training set and used to predict those in the test set.

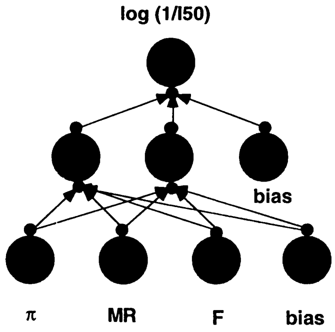
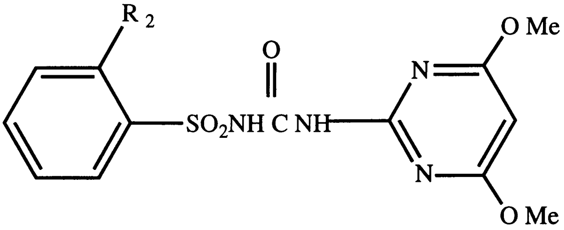


Figure 1
Back propagation neural network architecture for correlating biolgical activity with physico-chemical properties

TABLE 1 Structures and log ($1/I_{50}$) for the Inhibition of Acetolactate Synthase Enzyme by Benzenesulfonylureas



	R_2	log $1/I_{50}$		R_2	log $1/I_{50}$
1	SC_2H_5	9.12	21	$N(CH_3)_2$	8.10
2	SC_3H_7	9.01	22	$SO_2C_2H_5$	8.01
3	$SeCH_3$	8.94	23	$CO_2C_3H_7$	8.00
4	I	8.82	24	$SO_2C_3H_7$	7.97
5	C_4H_9	8.75	25	$t\text{-}C_4H_9$	7.90
6	$n\text{-}C_3H_7$	8.74	26	SO_2CH_3	7.89
7	CO_2CH_3	8.64	27	C_6H_5	7.89
8	C_2H_5	8.61	28	CH_3	7.74
9	CH_2Cl	8.59	29	OCH_3	7.70
10	SCH_3	8.58	30	$NHCO_2C_2H_5$	7.57
11	NO_2	8.54	31	NH_2	7.22
12	$SCHF_2$	8.51	32	OH	7.17
13	SCF_3	8.50	33	$SO_2\ i\,C_3H_7$	7.15
14	$CO_2C_2H_5$	8.47	34	F	7.07
15	$SO_2N(CH_3)_2$	8.42	35	$CH_2C_6H_5$	6.74
16	$i\text{-}C_3H_7$	8.38	36	$P(O)(OCH_3)_2$	6.68
17	Br	8.37	37	H	6.50
18	Cl	8.34	38	$SO_2C_6H_5$	6.04
19	OCH_2CH_3	8.26	39	$P(O)(OC_2H_5)_2$	5.80
20	CH_2OCH_3	8.14			

In order to elucidate the relative significance of the each independent variable in explaining enzyme inhibitory activity as determined by neural networks, models were calculated for all 15 combinations of one, two, three, and four independent variables selected from π, MR, $\mathcal{F}$, and $\mathcal{R}$. For each combination, five neural models, corresponding to 1-5 hidden units, were calculated, generating a total of 75 models. The one with minimal test set variance and maximal squared correlation coefficient was selected and will be discussed below.

Results

Stepwise regression indicates the following order of significance of the independent variables: MR > π > $\mathcal{F}$. $\mathcal{R}$ was found to be insignificant. ALS inhibition is expressed by

$$\log 1/I_{50} = 6.554 + 0.222\,\mathrm{MR} - 0.0079\,\mathrm{MR}^2 + 0.385\,\pi + 0.830\,\mathcal{F}$$
$$N = 39\ ;\ S = 0.381\ ;\ R = 0.898$$
$$\mathrm{MR}_{\mathrm{opt}} = 14.05 \qquad (2)$$

Figure 2 is a graph of this equation over the (MR,π) subdomain. In equation 2, ALS inhibition increases with size for small functional groups, reaches a maximum at MR=14.05, and decreases for larger groups. Closer analysis of this parabola and its curvature indicates that optimal groups have 4-7 heavy atoms selected from carbon, oxygen, nitrogen, sulfur, phosphorus. Examples include CO_2CH_3, $CO_2C_2H_5$, $COCH_3$, $SO_2C_2H_5$, $CON(CH_3)_2$, etc. equation 2 also indicates that ALS inhibition increases with hydrophobicity. The dependence on MR and π suggests that R_2 binds in a hydrophobic pocket that accommodates 4-7 heavy atoms (3). The positive coefficient of $\mathcal{F}$ indicates that enzyme inhibition is higher for inductively electron withdrawing substituents.

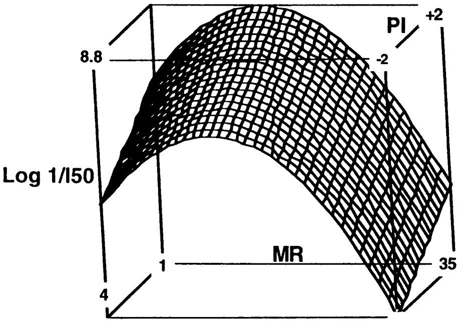


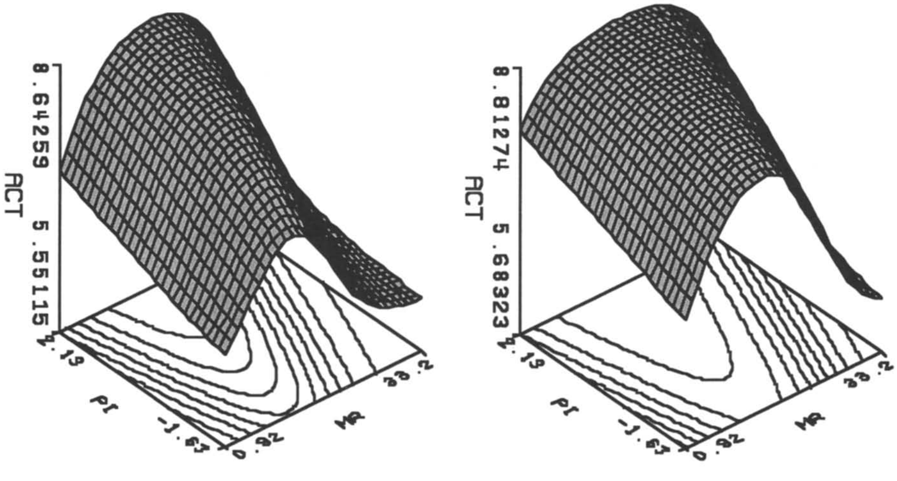
Figure 2
Graph of equation 2 over the (MR, π) subdomain

Furthermore, the curvature of the MR parabola in equation 2

$$\left(\frac{\partial^2 \log(1/I_{50})}{\partial\,\mathrm{MR}^2}\right)_{\pi,F} = -0.0158 \qquad (3)$$

is a constant, independent of π and $\mathcal{F}$. Its value is the same at the maximum of the parabola and throughout the surface. This curvature may be interpreted as tolerance to size variation: lower (less negative) curvature corresponds to more tolerance to size variation and *vice versa*. More succinctly, MR parabolas with lower curvature at the maximum retain high levels of biological activity over a wider range of size variation than parabolas with higher curvatures at the maximum. The constancy of this curvature suggests that R_2 fits in an enzyme pocket with fixed size.

Equation 2 is a typical regression based QSAR result in which activity depends linearly or parabolically on physico-chemical properties. Cubic or higher order terms typically are not used. Cross-products which correspond to coupling interactions between physico-chemical properties are prohibitively difficult to delineate. Due to the absence of cross-products in equation 2, optimal size is independent of hydrophobicity and/or electronic character.



(A) $\mathcal{F} = -0.08$ (B) $\mathcal{F} = 0.67$

Figure 3

Neural networks calculated biological activities (ACT) as a function of MR and π for a slightly electron donating R_2 (A) with $\mathcal{F} = -0.08$ and an electron withdrawing R_2 (B) with $\mathcal{F} = 0.67$

The neural networks based order of significance of independent variables is: MR > π > $\mathcal{F}$. $\mathcal{R}$ is insignificant. This matches the order determined with regression based QSAR. The (R^2, standard deviation) values are (0.834,0.345) and (0.894,0.216) for the training and tests respectively.

While neural models are expressed in closed analytical forms, unfortunately, due to the highly non-linear nature of these functions, they are not readily interpretable by inspection as is equation 2. In the current model however, ALS enzyme inhibition is a function of only three physico-chemical properties. In this simple case, it is possible to understand the qualitative and semi-quantitative behavior of the model by graphing the model calculated log ($1/I_{50}$) as a function of π, MR, at two or more values of $\mathcal{F}$ (Figure 3).

Figure 3A shows the model calculated log ($1/I_{50}$) (ACT) as a function of MR and π for a slightly electron donating group ($\mathcal{F} = -0.08$). While this surface is not parabolic in a strict mathematical sense, it has a "parabolic or Guassian-like" shape, with a ridge of maxima along the MR axis. Close inspection shows that optimal MR is approximately 10 for hydrophilic substituents (π = -1.63), and 15 for hydrophobic ones (π = 2.13), suggesting that tolerance of the binding pocket for size increases with hydrophobicity: the binding pocket is able to tolerate larger hydrophobic substituents than hydrophilic ons. Moreover, the curvature at the maximum is smaller for hydrophobic functionalities than for hydrophillic ones indicating that higher activity is retained for broader size variations of hydrophobic substituents than hydrophilic ones.

Figure 3B is the corresponding surface for an electron withdrawing substituent ($\mathcal{F} = 0.67$). Although it is similarly shaped to Figure 3A, close comparisons indicates that, in general, curvature of the surface in 3B is less than that in 3A, suggesting that the binding pocket tolerates wider size variations in electron withdrawing substituents than electron donating ones. These insights are different from those obtained from regression based QSAR where optimal size and tolerance to size variation are independent of hydrophobicity or electronic character.

In conclusion, for the data set studied in this manuscript, neural networks and regression QSAR's identify the same order of significance of independent variables: $MR > \pi > \mathcal{F}$. $\mathcal{R}$ is insignificant. Although the general shapes of the structure-activity surfaces in the two methods are similar, some differences are readily discernable. In spite of their subtlety, these differences lead to intriguing and dissimilar insights, unique to each method. Regression QSAR suggests an inflexibly size limited binding pocket. By contrast, neural QSAR suggests a flexible bindig pocket with variable size, determined by the totality of the ligand's physico-chemical properties. These results are consistent with those obtained from x-ray crystallographic studies of protein/ligand complexes which show that even structurally related ligands can bind differently to the same protein.

Literature Cited

(1) Andrea, T. A. & Kalayeh, H. In *QSAR: Rational Approaches to the Design of Bioactive Compounds* ; Silipo, C. & Vittoria, A. Eds.; Elsevier; Amsterdam, 1991, 209-121.

(2) Andrea, T. A. & Kalayeh, H. *J. Med. Chem.*, 1991, 34, 2824.

(3) Aoyama, T., Suzuki, Y., & Ichikawa, H. *J. Med. Chem.*, 1990, 33, 2583.

(4) Levitt, G. *The Discovery of Sulfonylurea Herbicides* ; Baker, D., Feneyes, J., Moberg, W., Eds.; *In Synthesis and Chemistry of Agrochemicals, Part II;* American Chemical Society; Washington, D. C., 1991.

(5) Andrea, T. A., Artz, S. P., Ray, T. B., Pasteris, R. J. In *Structure-Activity Relationships in Sulfonylurea Herbicides*; Draber W. & Fujita T., Eds.; Rational Approaches to Structure, Activity, and Ecotoxicology of Agrochemicals; CRC Press; Boca Raton, Florida, 1992, 373-395.

(6) Hansch, C. and Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley; New York, 1979.

(7) Owens, A. J. and Filkin, D. L.; *Joint IEEE/INNS International Joint Conference of Neural Networks*; Washington D. C.; 1989, 381.

RECEIVED April 26, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

Chapter 22

Three-Dimensional Quantitative Structure–Activity Analysis of Steroidal and Dibenzoylhydrazine-Type Ecdysone Agonists

Yoshiaki Nakagawa[1], Bun-ichi Shimizu[1], Nobuhiro Oikawa[1], Miki Akamatsu[1], Keiichiro Nishimura[1], Norio Kurihara[2], Tamio Ueno[1], and Toshio Fujita[1,3]

[1]Department of Agricultural Chemistry and [2]Radioisotope Research Center, Kyoto University, Kyoto 606–01, Japan

Comparative molecular field analysis (CoMFA) was utilized to examine the relationship between structure and insect molting hormone activity of substituted *N-t*-butyl-*N,N'*-dibenzoylhydrazines. The CoMFA features well reproduced the structure-activity inferences deduced from a traditional QSAR previously studied. A set of compounds combining dibenzoylhydrazines and structurally different steroidal molting hormones such as 20-hydroxyecdysone was then submitted to the CoMFA. Among a number of superimposition procedures with which the molecular field descriptors are evaluated, the most probable candidates to give reasonable qualities for the structure-activity relationship were selected. This study is regarded to be a preliminary work from which to look for the next direction for the syntheses of better molting-hormone-active compounds.

Steroidal molting hormone, 20-hydroxyecdysone (I), regulates insect metamorphoses. However, the external administration of an excessive amount of the molting hormone-active compounds causes the premature molting of larvae leading to their death (*1*). Recently, structurally very different dibenzoylhydrazines (II) have been disclosed to be agonists of the molting hormone (*2, 3*), and tebufenozide (II: X = 3,5-Me_2, Y = 4-Et) has been commercialized as a selective insecticide, especially against lepidopteran pest insect larvae (*4*).

We have measured the potency of these two types of compounds to promote the incorporation of *N*-acetylglucosamine into cultured integument fragments from the rice stem borers, *Chilo suppresssalis* Walker (*5-7*). For 20-hydroxyecdysone and its steroidal analogs, the potency measured by this procedure has been shown to parallel that measured in terms of the evagination of imaginal discs of *Drosophila melanogaster* (*7*), a conventional procedure to estimate the insect molting hormone activity (*8*). Because these two types of compounds are believed to share common receptor sites from experimental observations using the receptor preparation from the

---

[3]Current address: EMIL Project, Fujitsu Kansai Systems Laboratory, 2–2–6 Shiromi, Chuo-ku, Osaka 540, Japan

0097–6156/95/0606–0288$12.00/0

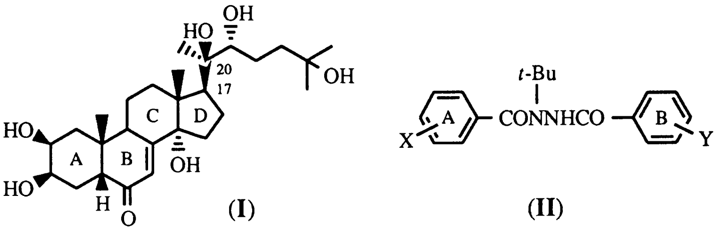


Figure 1. 20-Hydroxyecdysone (I) and Dibenzoylhydrazines (II).

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

Drosophila Kc cells (*3*), we have been interested in elucidating the structural factors required for the molting hormone activity which are common to two series of compounds.

Previously, we have analyzed quantitatively the index of the molting hormone activity measured using the rice stem borer integument system for a number of dibenzoylhydrazines carrying substituents on the two benzene rings in terms of physicochemical substituent effects (with the traditional QSAR procedure) (*6*). We have also examined qualitatively the structure-activity relationship of 20-hydroxyecdysone and its analogs, in which the structure of the side chain at position 17 of the steroidal skeleton is varied (*7*). Because these two sets of compounds belong to "entirely" different structural series, it is impossible to combine them as a single set for the traditional QSAR. Moreover, the number of compounds included in the two series is not well balanced. Some forty "systematically" substituted compounds are included in the dibenzoylhydrazines, whereas there are only five steroidal hormones the potency value of which has been measurable.

In this chapter, procedures which we took under these circumstances for the structure-activity analyses will be described. We used the comparative molecular field analysis (CoMFA) (*9*), a variation of the QSAR which enables us to deal with a combined set of structurally different series of compounds. First, we applied the CoMFA procedure to the set of dibenzoylhydrazines. We compared the CoMFA result with that previously obtained from the traditional QSAR, and examined as to whether they are conformable with each other. Then, we attempted to select superimposition procedures of the skeletal structure of dibenzoylhydrazines and the steroidal structure of 20-hydroxyecdysone and its analogs on the basis of intuitive and three-dimensional structural "similarities" as well as the structure-activity information for each series of compounds. The most probable superimposition procedures would be those with which the structural requirements of the combined set of compounds for the molting hormone activity could be elucidated by common patterns of molecular fields surrounding the entire molecule three-dimensionally. Because the CoMFA procedure was not practical to apply to only five steroidal ecdysone agonists, we examined how far the patterns of the molecular fields for dibenzoylhydrazines alone were affected after the steroidal compounds were superimposed. The less affected the molecular field patterns after the superimposition of steroidal agonists, the more "reliable" would be the prediction of the molting hormone activity of dibenzoylhydrazines and, in turn, that of the combined set of compounds including steroidal ecdysones.

Methods

Compounds. Dibenzoylhydrazines (**II**) and 20-hydroxyecdysone (**I**) and their analogs dealt with in this article are shown in Table I. Substituted dibenzoylhydrazines are classified into two groups, in one of which the monosubstitution is made on the A-ring and in the other of which the A-ring substituent is fixed as the 2-Cl but the B-ring is mono-substituted. The structure of steroidal molting hormones is indicated in Table II. In addition, we have synthesized a compound in which the carbonyl group adjacent to the *t*-Bu-amino moiety in

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

Table I. Molting Hormone Activity of Dibenzoylhydrazines and Ecdysones

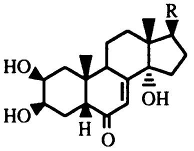
Structure II			pEC_{50} (M)						
					Correlation number[c]				
No.	X	Y	Obsd.[a]	Eq.1[b]	6	9	10	11	log *P*[d]
1	H	H	6.40	6.77	6.48	6.58	6.47	6.52	2.45
2	2-Cl	H	7.51	7.49	6.90	6.92	6.93	7.01	2.59
3	2-NO$_2$	H	6.80	7.04	6.74	7.06	6.90	6.88	2.27
4	2-CH$_3$	H	6.49	6.47	6.82	6.65	6.73	6.77	2.75
5	2-C$_6$H$_5$	H	<5.68	5.65	7.19	7.53	7.32	7.23	3.60
6	2-OCH$_3$	H	6.50	6.51	6.71	6.80	6.77	6.80	2.04
7	3-F	H	6.52	6.67	6.33	6.52	6.43	6.60	2.78
8	3-Cl	H	7.19	6.85	7.01	6.99	6.94	6.83	3.28
9	3-CF$_3$	H	6.42	6.64	6.28	6.24	6.43	6.45	3.61
10	3-NO$_2$	H	5.64	5.83	5.58	5.61	5.69	5.43	2.73
11	3-CN	H	5.42	5.61	5.72	5.62	5.60	5.70	2.34
12	3-CH$_3$	H	7.04	6.64	6.89	6.94	6.93	6.80	2.79
13	3-OCH$_3$	H	6.30	6.16	6.68	6.81	6.82	6.67	2.56
14	4-Cl	H	6.68	6.50	6.38	6.39	6.43	6.43	3.42
15	4-I	H	5.85	6.05	6.19	5.97	5.93	5.98	3.78
16	4-NO$_2$	H	5.12	4.79	4.97	5.18	5.25	5.15	2.63
17	4-CN	H	5.12	4.99	5.24	5.05	5.03	4.93	2.50
18	4-CH$_3$	H	6.15	6.21	5.99	5.95	5.89	5.98	2.99
19	4-OCH$_3$	H	5.21	5.48	5.25	5.23	5.41	5.22	2.56
20	2-Cl	2'-F	7.62	7.20	6.92	7.14	6.97	7.05	2.63
21	2-Cl	2'-Cl	6.62	6.70	6.60	6.60	6.78	6.83	2.75
22	2-Cl	2'-Br	6.61	6.54	6.64	6.77	6.79	6.85	2.91
23	2-Cl	2'-CF$_3$	5.89	5.97	6.06	6.02	5.90	6.12	3.02
24	2-Cl	2'-CH$_3$	7.09	6.68	7.00	7.04	6.83	6.96	2.91
25	2-Cl	2'-OCH$_3$	5.52	5.84	5.80	5.71	5.61	5.68	2.37
26	2-Cl	3'-Cl	7.20	7.19	7.00	7.04	7.06	7.08	3.49
27	2-Cl	3'-I	6.52	6.61	6.76	6.93	6.90	6.85	3.87
28	2-Cl	3'-CF$_3$	6.00	6.16	6.06	6.24	6.14	6.25	3.70
29	2-Cl	3'-NO$_2$	5.86	5.85	5.46	5.84	5.75	5.94	2.73
30	2-Cl	3'-CH$_3$	6.98	6.60	6.66	6.85	6.99	6.89	3.11
31	2-Cl	4'-Cl	7.05	7.58	7.38	7.27	7.21	7.20	3.51
32	2-Cl	4'-Br	7.37	7.53	7.41	7.24	7.20	7.19	3.73
33	2-Cl	4'-I	7.33	7.38	7.58	7.46	7.41	7.31	3.96
34	2-Cl	4'-CF$_3$	7.08	6.91	7.02	6.82	6.96	6.80	3.68
35	2-Cl	4'-NO$_2$	6.53	6.46	6.47	6.47	6.59	6.51	2.78
36	2-Cl	4'-CN	5.92	6.31	6.56	6.28	6.47	6.39	2.44
37	2-Cl	4'-CH$_3$	7.38	7.09	7.30	7.24	7.30	7.28	3.15
38	2-Cl	4'-OCH$_3$	6.84	6.49	6.95	6.79	6.87	6.88	2.82
39	Ponasterone A		7.53	e	e	7.43	7.45	7.50	0.49[g]
40	20-Hydroxyecdysone		6.75	e	e	6.64	6.54	6.70	-1.72[g]
41	Cyasterone		6.37	e	e	6.39	6.57	6.62	-1.97[g]
42	Inokosterone		6.18	e	e	6.36	6.34	6.03	-1.50[g]
43	Makisterone A		5.73	e	e	5.55	5.46	5.52	-1.32[g]
44	Ecdysone		<5.27	e	e	7.11	6.92	7.16	-0.36[g]
45	Benzyl derivative[f]		5.10	e	e	4.78	4.79	4.87	3.92[g]

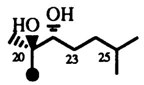
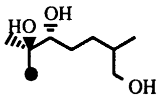
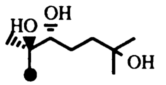
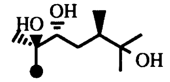
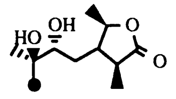
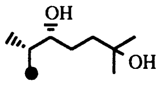
a) From refs. (*6*) and (*7*). b) Calculated by equation 1. c) Calculated by correlations in Tables III, IV, and V. d) From refs. (*12*) and (*13*). e) Not calculable. f) *N*-Benzyl-*N*-*t*-butyl-*N*'-benzoylhydrazine. g) Calculated by CLOGP (*14*).

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

compound **1** is reduced to methylene, the "benzyl derivative" (**45**) in Table I, and measured the activity with the rice stem borer integument system.

Table II. Structure of Ecdysone Analogs



No.	Name	R	No.	Name	R
39	Ponasterone A		**42**	Inokosterone	
40	20-Hydroxy-ecdysone		**43**	Makisterone A	
41	Cyasterone		**44**	Ecdysone	

Molecular Modeling. All computations were done with the molecular modeling software package SYBYL ver. 6.0 and 6.1. For dibenzoylhydrazines, the X-ray diffraction data of the unsubstituted (**II**: X = Y = H) and 2-Cl derivatives (**II**: X = 2-Cl, Y = H) were used to construct their initial conformation. These X-ray structures are depicted in Figure 2 showing that they do not differ much. The X-ray structure of the 2-Cl derivative was taken as the reference to build the initial conformation of substituted dibenzoylhydrazines because they are substituted at least on the A-ring. For steroidal ecdysones, the X-ray structure of 20-hydroxyecdysone shown in Figure 2 was taken as the reference (*10*). The coordinates of substructures to be modified in substituted derivatives were calculated with the SYBYL standard values for bond lengths and angles using the "Building" module.

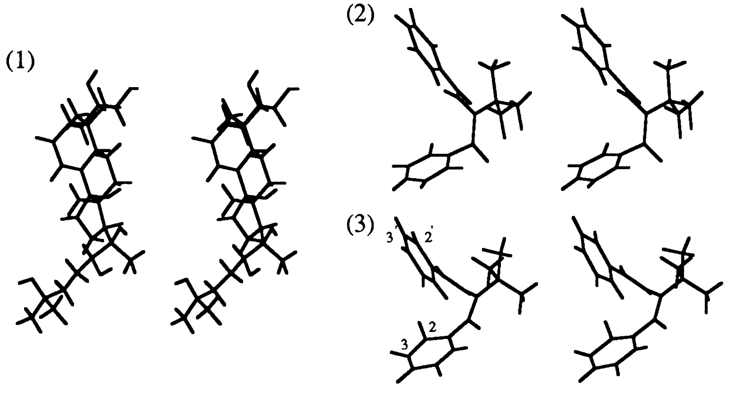


Figure 2. X-ray Structures of Reference Compounds. (1) 20-Hydroxy-ecdysone (**I**), (2) **II** : X = Y = H, (3) **II** : X = 2-Cl, Y = H.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

For dibenzoylhydrazines, the 2(*ortho*)-substituents other than the Cl-substituent were introduced into the same 2-position of the A-ring. The 3(*meta*)-substituents were placed into the 3-position. The fact that these two vicinal positions are less sterically hindered by other submolecular components than the positions on the other side of the C_1-C_4 axis is clearly observed in Figure 2. Likewise, the introduction of the 2(*ortho*)- and 3(*meta*)-substituents into the B-ring was made so that the substituents are located at the vicinal positions of the less congested side of the ring. The initial conformation of each compound was fully optimized by the semi-empirical molecular orbital method, PM3, in the program package MOPAC 5.0. The fully optimized conformation was regarded as the "active" structure of each compound.

CoMFA Procedure. The analyses were carried out using the QSAR option of SYBYL ver. 6.0 and 6.1. The lattice spacing was 2.0 Å and the +1 charge and the sp^3-carbon were used as probes to estimate the electrostatic and steric molecular field, respectively. The electrostatic and steric potential energies at each lattice point were calculated using the Coulombic and Lennard-Jones potential function, respectively. The atomic charges in each of the molecules were calculated using PM3.

To examine as to whether the hydrophobic molecular field exerted by individual submolecular components is significantly involved, the HINT software developed by Kellogg and coworkers (*11*) was operated along with the CoMFA. The addition of the hydrophobic lattice parameters was, however, not significant. The region-dependent hydrophobic effect was thought to be not involved. Thus, the participation of hydrophobicity was evaluated by using the molecular log *P* as a lattice-independent descriptor. The log *P* value for the dibenzoylhydrazines was either measured experimentally or estimated empirically (*12, 13*). For steroidal compounds, it was calculated using the CLOGP method (*14*). The CLOGP value (2.45) for the unsubstituted dibenzoylhydrazine (**1**) was identical with the value measured experimentally.

The correlation between the biological activity index and the log *P* and the lattice variables was analyzed by the partial least squares method. The CoMFA results were represented by the leave-one-out cross-validated correlation coefficient, *q*, and standard deviation, s_{press}, and the number of optimum latent variables, *m*, as well as the conventional correlation coefficient, *r*, and standard deviation, *s*, in addition to the weight percentage of the type of the descriptors involved in the correlation. The results were visualized by diagrams for contour lines connecting lattice points having an equivalent coefficient level for each molecular field surrounding a set of molecules.

Traditional QSAR Analysis and CoMFA of Dibenzoylhydrazines

Traditional QSAR. For 37 (un)substituted dibenzoylhydrazines in Table I, omitting the 2-C_6H_5 derivative (**5**) the activity of which is not accurately measurable, equation 1 has been formulated previously (*6*).

$$\begin{aligned}\mathrm{pEC_{50}} = {} & 0.98\ (\pm 0.26)\log P + 2.89\ (\pm 1.28)\ \sigma_I^{o(X)} - 0.78\ (\pm 0.49)\ \sigma^0_{o,m,p(X)} \\ & - 0.59\ (\pm 0.54)\ \Delta V_w^{o(X)} - 0.46\ (\pm 0.36)\ \Delta V_w^{m(X)} - 1.06\ (\pm 0.36)\ \Delta V_w^{p(X)} \\ & - 1.00\ (\pm 0.29)\ \Delta V_w^{o(Y)} - 1.25\ (\pm 0.31)\ \Delta V_w^{m(Y)} - 0.85\ (\pm 0.29)\Delta V_w^{p(Y)} \\ & + 4.38\ (\pm 0.77) \qquad n = 37 \quad s = 0.288 \quad r^2 = 0.872 \quad F_{9,27} = 20.43 \qquad (1)\end{aligned}$$

In equation 1, $\mathrm{pEC_{50}}$ is the log value of the reciprocal of the concentration $\mathrm{EC_{50}}$ (M) required to stimulate the *N*-acetylglucosamine incorporation into the rice stem borer integument system by 50% of the maximum (*5, 6*). The σ^0 is the parameter representing the total electron-withdrawing property for substituents devoid of the through-

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

resonance effect (*15*). The σ^0 value for *ortho* substituents is taken to be equivalent with that of the corresponding *para* substituents (*16*). The σ_I value (inductive component of electron withdrawing property) is also used for *ortho* substituents of the A-ring to correct the electronic effect simulated by σ^0_{para} (*17*). ΔV_w is Bondi's van der Waals volume (*18*) of substituents relative to that of hydrogen and scaled by 0.1 to make the size comparable to that of other parameters. The italicized super- and subscripts indicate the substitution position on the respective rings.

Equation 1 indicates that the activity is enhanced with the molecular hydrophobicity. The introduction of the electron-donating group into the A-ring is favorable to the activity in general. For the *ortho* substituent on the A-ring, however, the inductively electron-withdrawing effect outweighs the electron-donating effect. The introduction of the substituent at any position is unfavorable to the activity in terms of the negative ΔV_w. The unfavorable steric effect is lowest at the *meta* positions of the A-ring and at the *para* position of the B-ring among respective positional isomers. It should be noted that the structural features of the commercialized compound, tebufenozide (**II**: X = 3,5-Me_2, Y = 4-Et), are consistent with those for the potent compounds expected from the QSAR analysis.

CoMFA. Each "active" structure of 37 dibenzoylhydrazines was superimposed with that of the unsubstituted reference compound so that each structural component is as close as possible to the corresponding component of the reference. Thus, eight atoms, *i.e.*, the C_1 and C_4 atoms of both benzene rings and the four (C-N-N-C) bridging atoms, were selected and the sum of the squares of distance of the eight atomic positions from corresponding atomic positions of the unsubstituted compound was made as small as possible. The rotational "freedom" of the *t*-Bu group and the two benzene rings were not particularly considered, because the conformational patterns of each of these substructures were much the same in the active structures.

The CoMFA statistical results are shown in Table III. Reasonable results are not obtained without consideration of at least both the steric and electrostatic fields judged by the cross-validation statistical values in correlations 2 and 3. Even though the conventional correlation quality, r^2 and s, in correlation 3 appears to be high, it seems to be due to an artifact of using a greater number of latent variables. The additional use of the log P value as an independent variable improved the predictive ability in terms of the cross-validation even under conditions where the number of latent variables is unchanged as shown in correlations 4 and 5. If one more component is additionally used in correlation 6 corresponding to the additional use of log P, the conventional correlation in terms of r^2 and s becomes much better than that of correlation 4 although the cross-validation statistics are slightly poorer than those of correlation 5. We selected correlation 6 as the best considering both predictability and correlation ability. The improvement in the conventional correlation quality from correlation 5 to 6 was considered to surpass the impairment in the cross-validation quality.

Table III. CoMFA Correlation Statistics for Dibenzoylhydrazines (n = 37)

Descriptors used	Cross-validated		Conventional			Contribution			Cor.[c] No.
	q^2	s_{press}	r^2	s	m	st.[a]	el.[b]	log P	
st. only	0.174	0.653	0.563	0.475	2	100	-	-	(2)
el. only	0.161	0.689	0.916	0.218	5	-	100	-	(3)
st. + el.	0.339	0.593	0.795	0.330	3	57.4	42.6	-	(4)
st. + el. + log P	0.448	0.542	0.766	0.352	3	50.4	38.4	11.3	(5)
st. + el. + log P	0.438	0.555	0.845	0.292	4[d]	52.5	36.8	10.7	(6)

a) Steric field descriptors. b) Electrostatic field descriptors. c) Correlation number. d) Number of components was increased by one from that in correlation 5.

Views 1a and 1b are the contour diagram for the steric and electrostatic fields, respectively, with the 2-OCH_3 compound (6) inserted. In these and following contour diagrams, the contours are drawn on the ± 0.025 and ± 0.012 coefficient level for steric and electrostatic field, respectively. The negative steric contour is drawn with the yellow color representing the sterically forbidden region, whereas the positive contour with the green color is for the sterically accessible space. The negative electrostatic contour is drawn with the red color for the electronegative potential region, whereas the positive contour with the blue color is for the electropositive zone.*

View 1a indicates that, except for the 2(*ortho*)-position of the A-ring, all substituent positions are sterically forbidden. The substituent positions were defined so that the *ortho*, *meta*, and *para* monosubstitutions occur at the 2-, 3- and 4-positions on each of the benzene rings. Therefore, the contour lines do not appear at locations covering the 5- and 6-positions. Around the 3(*meta*)-position of the A-ring and the 4(*para*)-position of the B-ring, the yellow contours are less significant than those around the other positions indicating that the sterically forbidden degree of substituent occupation is lower at these positions. The degree seems highest at the *para* position of the A-ring. The appearance of the sterically permissible region at the *ortho* position of the A-ring is because the very weakly active 2-C_6H_5 compound (5) is not included in the analysis. An optimum should exist for the steric demand for the *ortho* substitution on the A-ring. The fact that no contour is observed at regions surrounding the *t*-Bu group indicates that no conformational variation of this group exists relating to the activity variations, supporting the procedure of superimposition performed without selecting any atom from the *t*-Bu group.

View 1b shows that the electropositive contour is drawn at positions corresponding to the *meta* and *para* substituents on the A-ring as well as at positions surrounding the *meta* and *para* substituents on the B-ring. The electropositive contour lines corresponding to the substituents on the A-ring indicate that the substituents are electron-donating toward the benzene ring so that their edges tend to be electropositive. This observation conforms with the negative ρ value for the A-ring substituent effect in equation 1. The fact that no clear electropositive contour is observed in the vicinity of the *ortho* position and the small electronegative region located below the *ortho* substituent position could probably correspond to the inductively electron-withdrawing σ_I term in equation 1. The electropositive contours at regions considerably apart from and the electronegative contours inside the positive contours surrounding the B-ring substituents would show that their electron-donating effect on the ring is less significant than that of the A-ring substituents.

The above arguments indicate that the CoMFA procedure represents essentially the same inferences as those deduced from the traditional QSAR. In correlation 6, the coefficient of the log *P* term is 0.52 and the intercept is 4.73. No regiospecific hydrophobic molecular field was shown to be significantly involved in the structure-activity pattern as examined by the HINT program (results not shown). Although the slope of the log *P* term is about double in equation 1, a fraction of the log *P* value may be elucidated by electrostatic and steric properties. This could lead to a lower slope in correlation 6. The intecept value is close to that in equation 1.

CoMFA for the Combined Set for Dibenzoylhydrazines and Steroidal Ecdysones

Superimposition Procedures Depending Upon Two-Dimensional Similarity. Although the two series of structures are very different from each other, there is a similarity to a certain extent. If we compare the structural formula of *N*-*t*-Bu-dibenzoylhydrazine with that of 20-hydroxyecdysone two-dimensionally, there seems to be a correspondence of "key atoms" between the two molecules as indicated with circles in Figure 3. In addition to 36 substituted dibenzoylhydrazines, 5 steroidal

---

*See color plate on page 299 for View 1.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

ecdysone analogs were superimposed on the unsubstituted dibenzoylhydrazine (**1**) so that the sum of the squares of distance between corresponding key atoms is minimal. As the key atoms, we first selected 8 atoms marked by circles in Figure 3 (superimposition I-1). Secondly, from the 8 atom pairs, 2 pairs with the shadowed circle were omitted (superimposition I-2). We finally reduced the number of key-atom pairs to 4 enclosed by a "large" circle (superimposition I-3). The "benzyl derivative" (**45**) was also included in the analyses. The superimposition procedure of this compound **45** on compound **1** was the same as that used for the dibenzoylhydrazines. The CoMFA statistics are shown in Table IV.

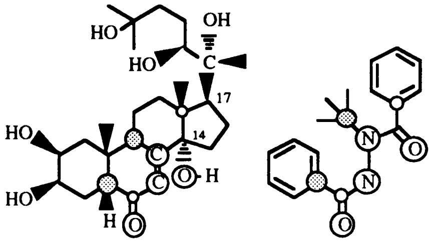


Figure 3. Two-Dimensional Similarity between 20-Hydroxy-ecdysone and *N-t*-Bu-dibenzoylhydrazine and the Key Atoms for their Superimposition.

The three-dimensional structures of dibenzoylhydrazines and steroidal ecdysone analogs, being similar to their X-ray structures shown in Figure 2, are by no means planar. Therefore, to use 8 atoms for the least squares method for the superimposition perhaps worsened the superimposition patterns of the COC=CC-O moiety of steroids on the CO-N-N-CO structure of the dibenzoylhydrazine. Thus, the cross-validated as well as conventional correlation statistics for superimposition I-1 are poor. With the use of only 4 atoms included in the substructure, COC=CC-O, of steroids and corresponding atoms in the dibenzoylhydrazine, the statistics are much improved as in superimposition I-3. Views 2a and 2b are the steric and electrostatic contour diagram, respectively, for superimposition I-3.* For the sake of simplicity, the structure of steroidal ecdysone is not inserted in these Views. The superimposition of the unsubstituted dibenzoylhydrazine and 20-hydroxyecdysone is shown separately in Figure 4. Such elements as the steroidal C-ring and a part of the B-ring are the only substructures to be superimposed on those of the dibenzoylhydrazine.

Table IV. CoMFA Correlation Statistics for the Superimpositions I for the Combined Set of Compounds (n = 43)

		Cross-validated		Conventional			Contribution			Cor.[c]
No.[a]	Conditions[b]	q^2	s_{press}	r^2	s	m	st.	el.	log P	No.
I-1	8 Atoms	0.146	0.682	0.717	0.393	3	47.1	40.8	12.1	(7)
I-2	6 Atoms	0.320	0.617	0.831	0.307	4	49.6	34.9	15.6	(8)
I-3	4 Atoms	0.393	0.590	0.900	0.240	5	46.4	40.5	13.2	(9)

a) Superimposition number. b) Number of atoms used for the least squares superimposition. c) Correlation number.

There are other procedures in which the carbonyl group in the steroidal ecdysones is superimposed on one of the carbonyl groups of dibenzoylhydrazines. However, such a superimposition procedure as that in which the other carbonyl

---

*See color plate on page 299 for View 2.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

oxygen of the dibenzoylhydrazines is superimposed on the 3-OH oxygen of steroids showed only poorer statistics (not shown).

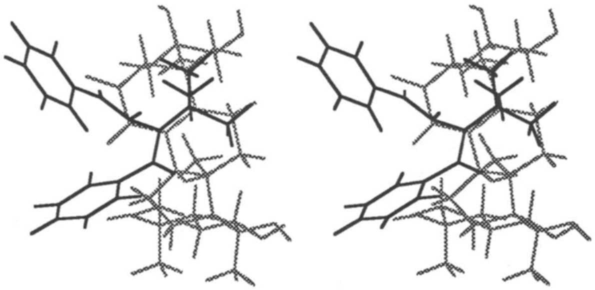


Figure 4. Stereoview for Superimposition I-3 between *N*-*t*-Bu-*N*,*N*'-Dibenzoylhydrazine (**1**) and 20-Hydroxyecdysone (**40**).

Superimposition Procedures Considering Three-Dimensional Structural Features. The most conspicuous three-dimensional structural feature shared by the two series of compounds is that both are of the bending type. As shown in Figure 2, two structural parts of the dibenzoylhydrazines are bending sharply with a certain angle at one of the bridge nitrogen atoms carrying the *t*-Bu group. In the 20-hydroxyecdysone molecule, the steroidal B-, C-, and D-rings are nearly "coplanar". The side chain of the position 17 is bent at the position 20, and extended holding a nearly "planar" structure making a certain angle with the steroidal skeleton. Thus, the superimposition of two sets of compounds could be made on the basis of the bending geometry.

For the superimposition of the bending "sites", the tertiary C atom of the *t*-Bu group was selected for the dibenzoylhydrazines. The molecules are actually bending at the N atom to which the *t*-Bu group is attached. If this N-atom was selected, however, the *t*-Bu group is protruded from the bending domain and does not overlap well with structural components of steroidal ecdysones. As the bending "site" of ecdysone analogs, we took the C atom at position 17 instead of that at position 20 where the side chain is actually bent. If the 20-C atom is selected, the distance from the OH oxygen attached to this C atom is too short compared with that between the tertiary C and the carbonyl oxygen in dibenzoylhydrazines. The 20-OH oxygen in ecdysones is to be superimposed on the carbonyl oxygen in dibenzoylhydrazines. As shown in Tables I and II for the structure-activity pattern of steroidal ecdysones, the OH group on the 20-C atom is essential. The OH group on the 25-C atom is, however, not important because ponasterone A (**39**) lacking this group is very active. The two OH groups attached to the 14-C and 20-C atoms were considered to be necessary to enhance the molting hormone activity of the steroidal ecdysones (*19*). For dibenzoylhydrazines, the carbonyl group next to the *t*-Bu-amino moiety is necessary for the high activity, because the benzyl derivative (**45**) was only very weakly active (see Table 1). Because both carbonyl groups are required in dibenzoylhydrazines (*2*), we considered that these two carbonyl oxygens correspond to two hydroxy oxygens in the steroidal ecdysones.

Furthermore, the C atom at the 25-position of 20-hydroxyecdysone was taken to be superimposable on the C atom at the *para* position of either the A- or B-ring of the dibenzoylhydrazines. The activity of dibenzoylhydrazines decreases with the bulkiness of substituents at both *para* positions. The steric congestion around the 25-C atom in steroidal ecdysones tends to be unfavorable to the activity, as observed in Tables I and II. For cyasterone (**41**), the C atom next to the lactone-carbonyl was chosen as the superimposable one.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

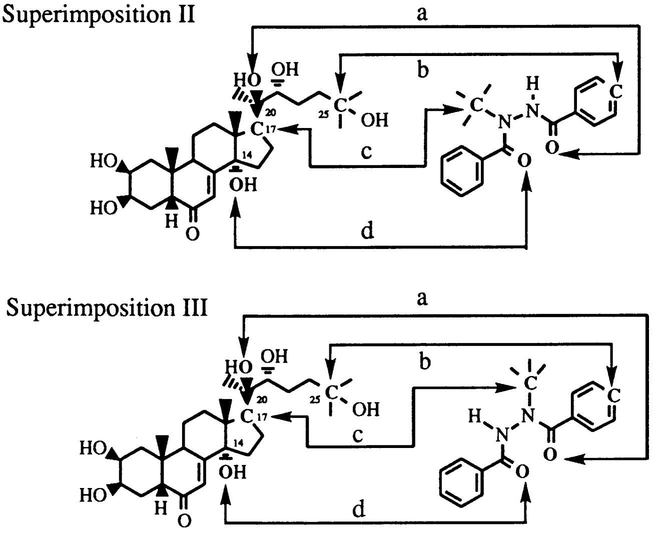


Figure 5. Superimpositions between 20-Hydroxyecdysone and *N*-*t*-Bu-Dibenzoylhydrazine Referring to Three-Dimensional Bending Structures. "a", "b", "c", and "d" indicate atomic pairs to be superimposed.

Considering how to superimpose two OH oxygen atoms of steroidal ecdysones on the two carbonyl oxygen atoms of dibenzoylhydrazines, two procedures, superimpositions II and III, referring to their bending structures, are illustrated two-dimensionally in Figure 5. In the least squares superimposition procedures, the weight of 3 was attached to the pairs "a" and "c" among four pairs. We considered that the bending geometry and the nearby oxygen atom are more important than other possibilities in the superimposition. The CoMFA statistics are shown in Table V. Unless the contribution of log *P* is considered, the statistics are much poorer (not shown). The Views 3a and 4a are the steric and 3b and 4b are the electrostatic contour diagrams for the superimposition II and III, respectively.* The 2-CH_3O compound (**6**) is inserted in each diagram. The least squares superimpositions of the unsubstituted dibenzoylhydrazine and 20-hydroxyecdysone according to the procedures II and III are in Figures 6 and 7, respectively.

Table V. CoMFA Correlation Statistics for the Superimpositions II and III for the Combined Set of Compounds (n = 43)[a]

	Cross-validated		Conventional			Contribution			Cor.[c]
No.[b]	q^2	s_{press}	r^2	s	m	st.	el.	log P	No.
II	0.431	0.571	0.883	0.259	5	48.0	39.8	12.2	(10)
III	0.472	0.551	0.892	0.250	5	49.8	40.7	9.5	(11)

a) The CoMFA region was taken as the same as that used in correlations 7 - 9 in Table IV. b) and c) See footnotes of Table IV.

*See color plate on page 299 for Views 3 and 4.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

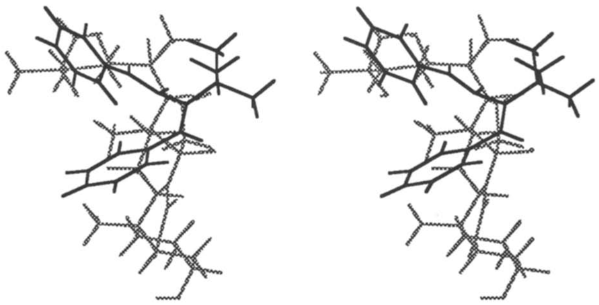


Figure 6. Stereoview for Superimposition II between *N*-*t*-Bu-*N*,*N*′-Dibenzoylhydrazine and 20-Hydroxyecdysone.

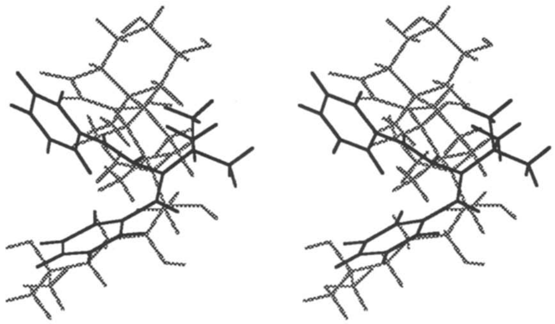


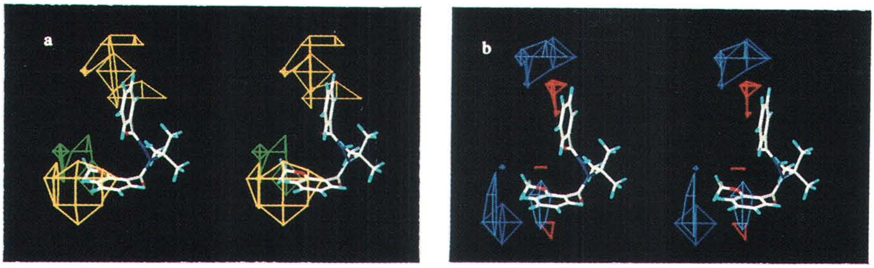
Figure 7. Stereoview for Superimposition III between *N*-*t*-Bu-*N*,*N*′-Dibenzoylhydrazine and 20-Hydroxyecdysone.

Comparison of the CoMFA results. Superimpositions II and III in Table V use 5 components as does superimposition I-3 in Table IV. The comparison among them is meaningful. While the cross-validation statistics are better in superimpositions II and III than those in I-3, the conventional correlation quality is almost identical among them.

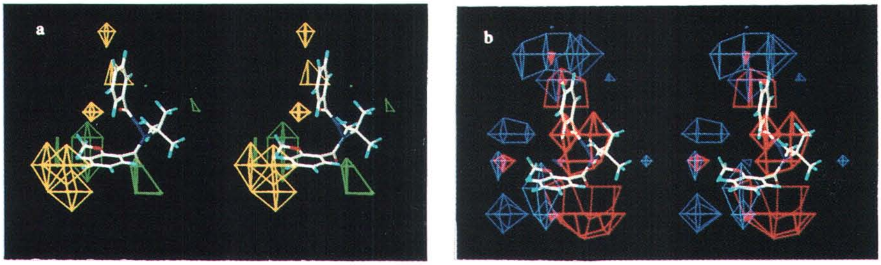
The similarities in steric and electrostatic contour diagrams for superimpositions I-3, II, and III to corresponding diagrams for correlation 6 in Table III for (un)substituted dibenzoylhydrazines were examined. It was not easy to judge as to how far the respective molecular field contours shown in Views 2 - 4 are "varied" from the corresponding references shown in View 1. Thus, the comparisons are inevitably qualitative. In Views 2b, 3b, and 4b, a large electronegative contours appeared at a region surrounding the carbonyl next to the *t*-Bu-amino moiety of the dibenzoylhydrazines. This reflects the importance of the electron-withdrawing carbonyl group for the high activity, because the weakly active benzyl derivative (**45**) is included in the analyses.

Because each of the superimposition procedures I-3, II, and III includes 37 dibenzoylhydrazines, similarities in the contour diagrams should exist to a certain extent at the vicinity of the dibenzoylhydrazine region. In View 2a for the superimposition I-3, however, the sterically accessible region is observed close to the 6(*ortho*)-position of the A-ring of the dibenzoylhydrazines, which is supposed to cover the region for the side chain of the steroidal compounds. In View 2b, electronegative contours emerged around the bending region. Moreover, because the major part of the side chain and the A- and B-rings of 20-hydroxyecdysone do not

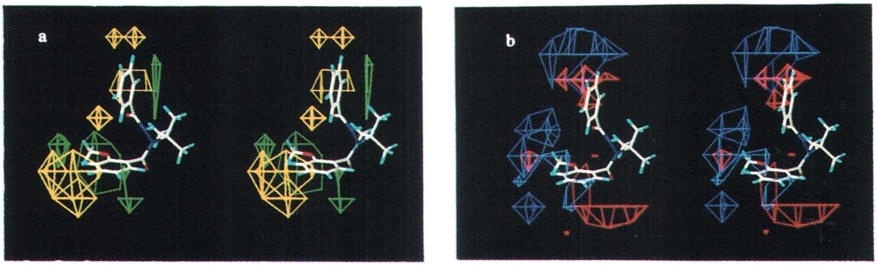
Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022



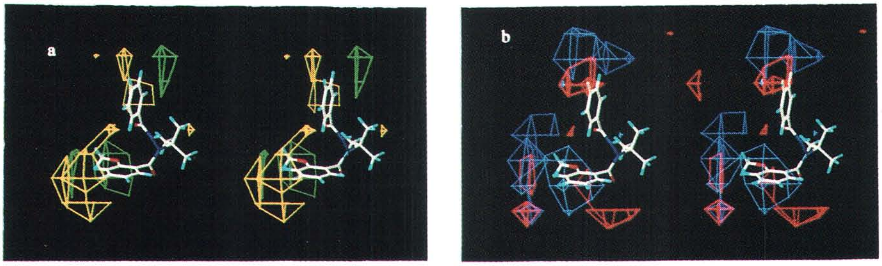
View 1. Stereoview of Contour Diagrams of the Steric (a) and Electrostatic (b) Fields for Correlation 6 in Table III.



View 2. Stereoview of Contour Diagrams of the Steric (a) and Electrostatic (b) Fields for Correlation 9 in Table IV (Superimposition I-3).



View 3. Stereoview of Contour Diagrams of the Steric (a) and Electrostatic (b) Fields for Correlation 10 in Table V (Superimposition II).



View 4. Stereoview of Contour Diagrams of the Steric (a) and Electrostatic (b) Fields for Correlation 11 in Table V (Superimposition III).

superimpose well with the dibenzoylhydrazine substructures in superimposition I-3 as shown in Figure 4, superimposition I-3 was not selected.

Of superimpositions II and III, we preferred to select the latter. The similarity to the contour diagrams in Views 1a and b for dibenzoylhydrazines seems to be higher in superimposition III in Views 4a and b than in superimposition II in Views 3a and b. The superimposition III seems to conform with the structure-activity information fed into the assumption to set up the superimposition procedure better than the superimposition II. For instance, the steric restriction occurring at the *para* position of the A-ring in dibenzoylhydrazines appears to apply to that at position 25, the end of the side chain in steroidal ecdysones better than that of the B-ring in superimposition II. The comparison between Figures 6 and 7 indicates that most of the structural components, except for the A-ring, of steroidal ecdysones fit better into those of the dibenzoylhydrazines in superimposition III than in superimposition II.

Discussion

Although superimposition III was selected here as the most acceptable procedure for the set of dibenzoylhydrazines and steroidal ecdysones, there could be others which exceed this in quality for the present combination of compound sets. As far as the CoMFA correlation statistics are concerned, the superimpositions are almost equivalent among the procedures I-3, II, and III. Even though the geometrical extent of superimposition seems so different among procedures as shown in Figures 4, 6, and 7, the CoMFA correlation quality was not much changed.

In the steroidal ecdysones, the structural change is only made on the side chain, while in the dibenzoylhydrazines, it is on both of the benzene rings. Even though the structural modifications are made in a set of compounds, structure-activity considerations in terms of structural features of a certain region would be impossible, unless modifications are made in this particular region. In superimposition I-3, the overlapping substructures, the bridge CONNCO moiety of dibenzoylhydrazines and the junction between B- and C-rings in steroidal ecdysones, are not changed in the respective series of compounds. In superimpositions II and III, one of the benzene rings of dibenzoylhydrazines overlaps the structurally unchanged skeleton of the ecdysone series of compounds. The CoMFA results, showing that correlations of high quality were obtained even with such superimposition procedures, might suggest that no matter how the steroidal ecdysones were superimposed on the dibenzoylhydrazines, the overall correlation results could be determined mostly by the CoMFA characteristics of the 37 dibenzoylhydrazines. The fact that there are cases showing very poor correlation statistics such as procedure I-1 in Table IV and one of the procedures in which the carbonyl oxygen and the 3-hydroxy oxygen of steroidal ecdysones are superimposed on the two carbonyl oxygens of the dibenzoylhydrazines ($q^2 = 0.256$, $s_{\mathrm{press}} = 0.592$, $m = 3$) suggests, however, that the superimposition procedures I-3, II, and III are by no means dominated by the CoMFA features for the dibenzoylhydrazines alone. The CoMFA correlation procedure utilizing molecular field parameters may account for long-range potentials at positions considerably apart from the substructures where the structural features such as substituents are variously modified.

The fact that the activity of the benzyl derivative (**45**) is well predicted by any of the procedures I-3, II, and III is a kind of support of the superimpositions. The failure in predicting the activity of ecdysone (**44**) is believed to be due to the fact that this compound, being inactive and the only one lacking the 20-OH group, is not included in the analyses. Although the 20-OH group is essential, the effect of substituent variations at the 20-position has not been well incorporated in the molecular fields. Further structural modifications should be made in the bending domain in steroidal ecdysone series to examine the structure-activity pattern relating to the side chain. It has been indicated that at least one of the benzene rings in

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch022

dibenzoylhydrazines need not be aromatic for the insecticidal activity (*2*). Benzyl and certain alkyl and alicyclic groups are replaceable without significant decrease in the activity. This justifies our procedure to superimpose the aliphatic side chain of the steroidal ecdysones on the substructure containing one of the benzene rings in dibenzoylhydrazines.

In conclusion, the present analyses still leave much to be clarified. It should be taken to be a reference study from which to start synthesizing new series of more potent molting hormone agonists in which "drastic" structural modifications are made according to the CoMFA results and those of the traditional QSAR. The structure-activity analyses to confirm the best superimposition model among candidates such as the procedures I-3, II, and III could be done on a trial and error basis to reach the final selection. It is emphasized that, in a single series of bioactive compounds such as substituted *N*-*t*-Bu-*N*,*N*'-diphenyhydrazines, the CoMFA procedure is able to reproduce structure-activity features deduced from the traditional QSAR when the structural variations are well designed.

Acknowledgments. We thank Drs. Tadamasa Da-te, Kimio Okamura, and Ryo Shimizu of Tanabe Seiyaku Co., Ltd. for the measurement of the X-ray crystal diffraction of dibenzoylhydrazines, and Dr. Hideo Abuki in Nippon Kayaku Co., Ltd. for the gift of ecdysone analogs.

Literature Cited

1) Kubo, I.; Klocke, J. A.; Asano, S. *J. Insect Physiol.* **1983**, *29*, 307-316.
2) Hsu, A. C-T. In *Synthesis and Chemistry of Agrochemicals II*; Baker, D. R.; Fenyes, J. G.; Moberg, W. K., Eds.; *ACS Symp. Ser. 443*; Amer. Chem. Soc.: Washington, D. C., **1991**; 478 - 490.
3) Wing, K. D. *Science* **1988**, *241*, 467-469.
4) Smagghe, G.; Degheele, D. *Pestic. Biochem. Physiol.* **1994**, *49*, 224-234.
5) Oikawa, N.; Nakagawa, Y.; Soya, Y.; Nishimura, K.; Kurihara, N.; Ueno, T.; Fujita, T. *Pestic. Biochem. Physiol.* **1993**, *47*, 165-170.
6) Nakagawa, Y.; Soya, Y.; Nakai, K.; Oikawa, N.; Nishimura, K.; Kurihara, N.; Ueno T.; Fujita, T. *Pestic. Sci.* **1995**, *43*, 339-345.
7) Nakagawa, Y.; Nishimura, K.; Oikawa, N.; Kurihara, N.; Ueno, T. *Steroids* **1995**, *60*, 401-405.
8) Fristrom, J. W.; Yund, M. A. In *Invertebrate Tissue Culture Research Application*; Maramorosch, K., Ed.; Academic Press: New York, NY, **1976**; 161-178.
9) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Amer. Chem. Soc.* **1988**, *110*, 5959-5967.
10) Huber, R.; Hoppe, W. *Chem. Ber.* **1965**, *98*, 2403-2424.
11) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comp.-Aided Molec. Design* **1994**, *5*, 545-552.
12) Oikawa, N.; Nakagawa, Y.; Nishimura, K.; Ueno, T.; Fujita, T. *Pestic. Sci.* **1994**, *41*, 139-148.
13) Oikawa, N.; Nakagawa, Y.; Nishimura, K.; Ueno T.; Fujita, T. *Pestic. Biochem. Physiol.* **1994**, *48*, 135-144.
14) Hansch, C.; Leo, A. J. CLOGP3 Release 3.54 Medicinal Chemistry Project, Pomona College: Claremont, CA, **1989**.
15) Fujita, T. *Prog. Phys. Org. Chem.* **1983**, *14*, 75-113.
16) Fujita, T; Nishioka, T. *Prog. Phys. Org. Chem.* **1976**, *12*, 49-89.
17) Charton, M. *Prog. Phys. Org. Chem.* **1981**, *13*, 119-251.
18) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441-451.
19) Nakanishi, K. *Steroids* **1992**, *57*, 649-657.

RECEIVED May 25, 1995

Chapter 23

Comparison of Classical QSAR and Comparative Molecular Field Analysis

Toward Lateral Validations

Ki Hwan Kim

**Department of Structural Biology, Abbott Laboratories,
Abbott Park, IL 60064**

Lateral validation refers to the method of validating new QSAR by laterally correlating it with various QSAR equations that support the new equation. This method was originally utilized in the classical QSAR field of physical organic chemistry and has recently been extended to the field of biological sciences. The possibility of supporting a new 3D-QSAR by lateral validation is investigated in this study using one of the most often used 3D-QSAR methodologies, Comparative Molecular Field Analysis (CoMFA).

Compared to the classical quantitative structure-activity relationship (QSAR), Comparative Molecular Field Analysis (CoMFA) offers several advantages in three-dimensional quantitative structure-activity relationship (3D-QSAR) *(1)*. For example, non-congeneric series can be studied, the information from the three-dimensional structure of molecules is included, and the issues of missing parameter values and the description of intermolecular steric effects in the classical QSAR with Taft Es parameters, which were derived from intramolecular steric effects, are no longer of concern. However, there are also several disadvantages in the CoMFA approach. Among them are the dependence on the accuracy of the superposition, the involvement of many adjustable parameters, and the lack of useful mathematical expressions of the results.

In classical QSAR, the results are presented in an equation that provides useful information in a concise manner. In the field of physical organic chemistry, the slopes and, to a lesser extent, the intercepts of QSAR equations are often compared, and such practices have now been extended to the field of biological QSAR. The terminology "lateral validation" has been coined for such applications *(2)*.

Lateral validation refers to the method of validating new QSAR by laterally correlating it with many other QSAR equations that support the new equation. Statistics alone are not enough to stay away from the murky cloud of chance correlations *(2)*, especially with biological QSAR where more complex biological effects in nonhomogeneous system are correlated with the structural changes. Even in the case where a QSAR standing alone means very little, lateral validation can support that the correlation was not obtained by a chance correlation and shed some light onto the equation. In order to aid lateral validations of QSAR, the Pomona Medicinal Chemistry Project has been compiling QSAR equations, and the raw data

0097–6156/95/0606–0302$12.00/0

supporting them, for many years. Currently, the data base includes about 6000 sets of data representing approximately 90,000 compounds. These equations are almost evenly divided between physical organic subjects and biomedicinal chemistry and toxicology subjects *(2)*.

Lateral Validation in Classical QSAR

Biological QSAR. The use of lateral validation in biological QSAR can be illustrated with the study of the role of the hydrophobicity in mutagenesis *(2)*. When equation 1 was developed for the mutagenicity of X-$C_6H_4CH_2N(Me)N=O$ compounds in *Salmonella typhimurium* TA1535, the slope of the hydrophobicity term of this equation was compared with eight other different QSAR from a variety of compounds acting in a number of different test systems. In equation 1, C is the molar concentration of compound producing a standard number of mutations in a fixed time interval. Excluding only one example, where a special colinearity problem occurs, the coefficients (h, slopes) of the hydrophobic term were found to be around 1 in each of these equations (See Table I). Thus, the similarity of these coefficients supported equation 1 by the lateral validation.

$$\log 1/C = 0.92\pi + 2.08\sigma - 3.26 \qquad n = 12,\ r = 0.891,\ s = 0.314 \qquad (1)$$

Table I. Coefficients (h) of Hydrophobicity in the QSAR of Bacterial Mutagenicity

Number of Compounds	Type of Compounds	Test	h
188	aromatic and heteroaromatic nitro compounds	TA98	0.65
67	aromatic and heteroaromatic amines	TA100	0.92
12	X-$C_6H_4CH_2N(Me)N=O$	TA1535	0.92
21	X-$C_6H_4N=NN(R)Me$	TA92	0.95
15	aromatic nitro compounds	*E. coli*	1.07
88	aromatic and heteroaromatic amines	TA98	1.08
117	aromatic and heteroaromatic nitro compounds	TA100	1.10
21	quinolines	TA100	1.14

Adapted from reference *(2)*.

Lateral comparisons of QSAR equations not only render support for the new QSAR, but also provide other valuable information. Comparing various QSAR equations with enzyme models constructed from X-ray crystallographic observations, Hansch gave new meaning to the coefficients *(3)*. He suggested that a coefficient approaching 0.5 indicates a substituent contacting an enzyme surface (or a receptor), and a coefficient approaching 1.0 indicates a substituent that has been engulfed by a hydrophobic pocket. The importance of optimum lipophilicity of around 2.0 expressed by the octanol-water partition coefficients for a variety of CNS agents is another product of lateral comparison *(4)*. Lateral comparisons of QSAR can also be made by comparing different biological systems interacting with similar compounds; many examples of lateral comparison of biological QSAR involving hydrophobic, steric, and/or electronic factors are described in a recent review *(2)*.

QSAR in Physical Organic Chemistry. Lateral validation has long been practiced in the field of physical organic chemistry. The electronic substituent constant σ, which was derived from the dissociation constants of substituted benzoic acids, and its several variations are used in thousands of correlations. The substituent constant σ measures the electronic effects of the substituent. In principle, it is independent of the nature of the chemical reaction, and its scale usually ranges from -1 to 1. On the other hand, the reaction constant ρ, which was defined to be 1.0 for the dissociation of benzoic acids in water at 25°C, measures the susceptibility of the reaction or equilibrium to the electronic substituent effects. It depends on the type and nature of the reaction, including conditions such as the solvent used and the reaction temperature. The unsigned values of the reaction constant often range from 0 to 4 *(5)*. Table II shows some examples. For the same type of reaction, such as the ionization of acids, the reaction constant varies significantly depending on the kind of side-chains involved even if the solvent and the temperature are the same. Despite such variations, the reaction constants often provide useful information. In Hansch's recent paper regarding QSAR in physical organic chemistry, *(2)* several examples addressing the importance of lateral validation were given. For example, the reaction constants of alkaline hydrolysis of aromatic carboxylic esters are approximately 2.0 even though the compounds involved and the reaction conditions were different (See equations 2-6) *(2)*. A comparison of these coefficients obtained from nonenzymatic systems was further extended to include some enzymatic hydrolyses of the same type of esters (See equations 7-8) *(2)*.

Table II. Dependence of the Reaction Constant (ρ) on Side-chain, Temperature, and Solvent for Reactions of Carboxylic Acids and Esters.

Substrate	Solvent	Temperature (°C)	Reaction constant (ρ)
Ionization of acid (same solvent and temperature)			
$ArCO_2H$	H_2O	25°	1.00
$ArCH_2CO_2H$	H_2O	25°	0.49
$ArCH_2CH_2CO_2H$	H_2O	25°	0.21
$ArCH{=}CH_2CO_2H$ (trans)	H_2O	25°	0.47
$ArSCH_2CO_2H$	H_2O	25°	0.30
$ArSO_2CH_2CO_2H$	H_2O	25°	0.25
Reaction of acid with diazodiphenylmethane (constant temperature)			
$ArCO_2H$	MeOH	30°	0.88
$ArCO_2H$	EtOH	30°	0.94
$ArCO_2H$	PrOH	30°	1.07
$ArCO_2H$	BuOH	30°	1.28
Reaction of acid with diazodiphenylmethane (same solvent and temperature)			
$ArCH_2CO_2H$	EtOH	30°	0.40
$ArCH_2CH_2CO_2H$	EtOH	30°	0.22
$ArCH{=}CH_2CO_2H$ (trans)	EtOH	30°	0.42
$p\text{-}ArC_6H_4CO_2H$	EtOH	30°	0.22
$ArOCH_2CO_2H$	EtOH	30°	0.25
Basic hydrolysis of ester (same solvent)			
$ArCO_2Et$	60% Me_2CO	0°	2.66
$ArCO_2Et$	60% Me_2CO	15°	2.53
$ArCO_2Et$	60% Me_2CO	25°	2.47
$ArCO_2Et$	60% Me_2CO	40°	2.38

Adapted from reference *(5)*.

Alkaline hydrolysis of X-C_6H_4-$COOCMe_3$ in 50% ethanol at 20°C

$\log k = 2.18\,\sigma + 0.62$ $n = 5$, $r = 0.997$, $s = 0.041$ (2)

Alkaline hydrolysis of X-C_6H_4-$COOC_2H_5$ in 87% ethanol at 30°C

$\log k = 2.51\,\sigma - 1.28$ $n = 18$, $r = 0.993$, $s = 0.105$ (3)

Alkaline hydrolysis of X-$C_{10}H_7$-1-$COOC_2H_5$ in 85% ethanol at 50°C

$\log k = 2.13\,\sigma - 2.58$ $n = 9$, $r = 0.998$, $s = 0.048$ (4)

Alkaline hydrolysis of X-pyridyl-2-COOMe in 85% methanol at 25°C

$\log k = 2.03\,\sigma - 2.09$ $n = 10$, $r = 0.997$, $s = 0.083$ (5)

Alkaline hydrolysis of X-C_6H_4-COOMe in aqueous solution

$\log k = 1.66\,\sigma + 1.92$ $n = 14$, $r = 0.999$, $s = 0.022$ (6)

Deacylation of X-C_6H_4-COO-chymotrypsin at pH 7.07 and 25°C

$\log k = 1.73\,\sigma - 2.07$ $n = 7$, $r = 0.971$, $s = 0.177$ (7)

Deacylation of X-C_6H_4-COO-chymotrypsin at pH 8.5

$\log k = 1.73\,\sigma - 3.48$ $n = 11$, $r = 0.960$, $s = 0.275$ (8)

Toward Lateral Validation in CoMFA

The analytical procedures of CoMFA can be divided into three steps: (1) representation of molecules by their steric, electrostatic, and sometimes hydrogen-bonding and hydrophobic interaction energies at the three-dimensional lattice points, (2) data analysis by the partial least squares (PLS) method, and (3) presentation of the results by 3D coefficient contour plots.

In the first step, various interaction energies are calculated employing different probe atoms or groups. The calculation of steric interaction energies is often accomplished with a CH_3 probe, the calculation of electrostatic interaction energies with an H^+ probe, and the calculation of hydrophobic interaction energies with a H_2O probe. In the data analysis step, thousands of interaction energy values calculated in the first step are transformed into a few principal components, often called latent variables and designated as Z1, Z2, Z3, ... in our CoMFA studies. These latent variables are then used in the regression analysis to derive a correlation equation.

A typical CoMFA paper fails to present the regression equation using the latent variables, but only reports the optimum number of components, number of compounds used, and correlation coefficients (R^2) and standard errors of estimate (RMSE or s) from the fitted model as well as the corresponding cross-validation test. In our studies, we have included the corresponding PLS regression equations in addition to the usual pertinent information. Since the coefficients provide valuable information in classical QSAR, it was our objective to gather similar information from the coefficients of PLS regression equations.

An interesting observation was made with the cases where the biological activity was related with log P in a "parabolic" or "bilinear" relationship *(6)*. The log MKC (minimum killing concentration) of benzyldimethylalkyl-ammonium chlorides against *Cl. welchii* serves an example. A two-component model was obtained from CoMFA. The plot of the two latent variables shows a "parabolic" or "nonlinear" shaped curve (Figure 1). Although it is not clear what one should interpret from such a plot at the present time, it appears that the latent variables contain potentially useful information.

Is it possible to extract useful information from the coefficients of PLS regression equations and incorporate that information into the lateral validation of new CoMFA correlations? This is the theme of this study. In order to address this important question, we studied the 3D-QSAR, using the CoMFA approach, of two relatively simple examples. The results are summarized below.

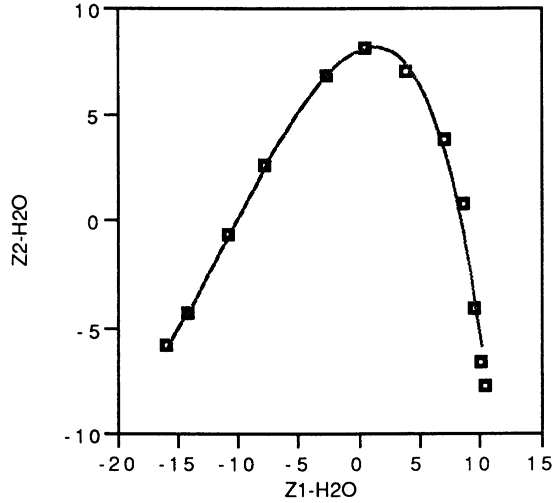
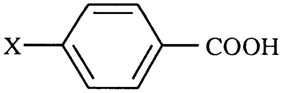


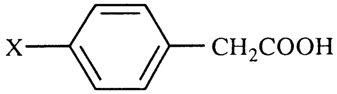
Figure 1. Plot of the first ($Z1_{H2O}$) and second ($Z2_{H2O}$) latent variables from CoMFA model (*Cl. welchii.*; Log MKC; Equation 8A of reference 6).

Dissociation Constants of Benzoic Acids and Phenylacetic Acids. In our previous publications *(7, 8)* we demonstrated that the electrostatic descriptors derived from an H^+ probe and the electrostatic potential using the CoMFA approach can describe the dissociation constants of forty-nine benzoic acids, twenty-three clonidine-like imidazolines and sixteen 2-substituted and 1-methyl-2-substituted imidazoles. In these cases, more than one latent variable was required in the CoMFA model due to the wide variety of substituent types included; the optimum number of components reflects the variations in the properties of interest and in the structures.

In this CoMFA analysis, only a small subset of the compounds is used in order to simplify the situation; the compounds used are H, p-Me, p-OMe, p-Cl, p-Br, and p-NO_2 derivatives of benzoic acid (BA, **I**) and phenylacetic acid (PhA, **II**). These are the para-substituents used by Shorter *(5)* which indicated that there was a high correlation between the ΔpK_a [= $pK_a(H)$ - $pK_a(X)$] values of the phenylacetic acids and the Hammett σ constant values derived from the benzoic acids (see equation 9). Equation 9 shows that the size of ΔpK_a values of phenylacetic acids is about half the size of the σ values based on the coefficient given by σ.



(I)



(II)

$$\Delta pK_a(PhA) = 0.507(\pm 0.048)\sigma + 0.032(\pm 0.018) \quad (9)$$
$$n = 6,\ r^2 = 0.966,\ s = 0.040,\ F = 112.0,\ p = 0.0005$$

The starting coordinates of the molecules were generated using the graphics modeling package for small molecules at Abbott. All geometric variables were optimized with AM1 of AMPAC. The molecules were aligned by superimposing the benzoic acid or phenylacetic acid moiety. The electrostatic potential energy fields of each molecule were calculated at the lattice points surrounding the molecules using an H^+ probe with the program GRID as described in our previous studies *(7, 8)*. For each molecule, the energies at a total of 1950 grid points were calculated with a 2Å spacing in a lattice of 28 x 24 x 18 (X = -10 to 18, Y = -11 to 13, Z = -9 to 9). All energy values greater than 10 kcal/mol were truncated to 10. Any lattice point for which the standard deviation of the energies was less than 0.05 was discarded. The effects of grid lattice position on the correlation were first investigated by shifting the grid box by -0.5, -1.0, and -1.5Å in each of the X, Y, and Z directions. Based on the results from these different grid box locations, the lattice position that was shifted from the original position by -0.5Å was chosen as the best location of the lattice points.

From the sets of benzoic acid and phenylacetic acid derivatives, excellent single component models were obtained by the CoMFA approach; equation 10 was from the six benzoic acid derivatives, and equation 11 was from the six phenylacetic acid derivatives. The quality of the correlation coefficients and the standard errors of estimate from both the fitted model and the leave-one-out cross-validation test are excellent. Equations 10 and 11 show that the latent variables Z1(BA) and Z1(PhA) account for 93% and 98% of the variances in the corresponding pK_a values, respectively. These latent variables were derived independently from the pK_a values and the three-dimensional structures of the corresponding benzoic acids or phenylacetic acids. Nevertheless, they show a remarkable similarity as shown in equation 12 and Figure 2; the slope of equation 12 is essentially 1, and the intercept is zero. The slope of 1 in equation 12 shows that there is a 1:1 relationship between Z1(PhA) and Z1(BA). Therefore, the size of the coefficient of Z1(PhA) in equation 11 is naturally about half of Z1(BA) in equation 10; this is consistent with the coefficient of σ described in equation 9.

$$pK_a(BA) = 0.043(\pm 0.006)Z1(BA) + 4.068(\pm 0.046) \quad (10)$$
$$n = 6,\ r^2 = 0.927,\ s = 0.114,$$
$$r^2_{cv} = 0.679,\ s_{cv} = 0.233,\ F = 50.7,\ p = 0.002$$

$$pK_a(PhA) = 0.023(\pm 0.002)Z1(PhA) + 4.212(\pm 0.012) \quad (11)$$
$$n = 6,\ r^2 = 0.982,\ s = 0.029,$$
$$r^2_{cv} = 0.953,\ s_{cv} =, 0.046\ F = 221.4,\ p = 0.0001$$

$$Z1(PhA) = 1.001(\pm 0.004)\ Z1(BA) + 0.000(\pm 0.033) \quad (12)$$
$$n = 6,\ r^2 = 0.9999,\ s = 0.081,\ F = 55596.1,\ p = 0.0000$$

The intercepts of equations 10 and 11 are not zero because pK_a values are used instead of ΔpK_a in these cases. Table III lists the pK_a values of benzoic acid and phenylacetic acid derivatives as well as the corresponding calculated values using equations 11 and 12.

The results obtained in this study indicate that the PLS latent variables Z1(BA) and Z1(PhA) provide similar information as the Hammett σ constant does at least in these single component models. Additionally, the results imply that it may be possible to use the coefficients of these descriptors for the purpose of lateral validation in CoMFA.

Results from the dissociation constants of benzoic acids and phenylacetic acids encouraged us to further investigate the possible use of the coefficients in a PLS regression equation; the rate constants of the elimination reaction from substituted phenylethylarenesulphonate analogs were used in the next example.

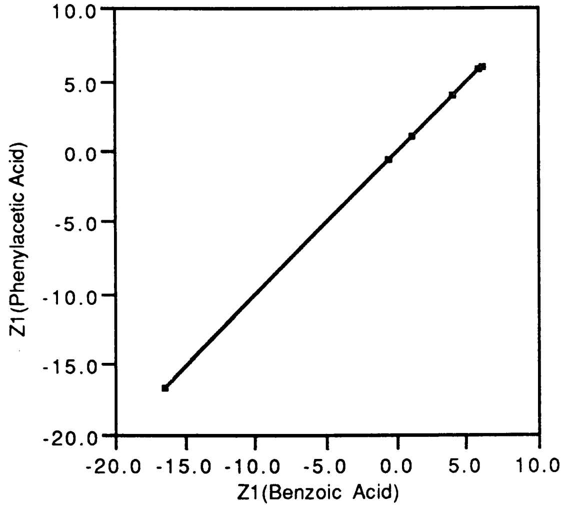


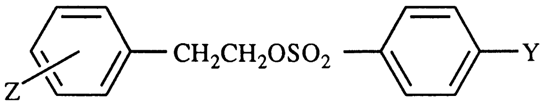
Figure 2. Plot between Z1(Benzoic Acids) versus Z1(Phenylacetic Acids).

Table III. Observed and Calculated pK_a Values of Benzoic Acids and Phenylacetic Acids Using Equations 10 and 11.

		pK_a (Benzoic Acids)			pK_a (Phenylacetic Acids)		
No	Substituent	obs	cal	dev	obs	cal	dev
1	H	4.20	4.24	-0.04	4.31	4.30	0.01
2	Br	3.97	4.12	-0.15	4.19	4.24	-0.05
3	Cl	3.98	4.04	-0.06	4.19	4.20	-0.01
4	Me	4.37	4.32	0.05	4.37	4.34	0.03
5	NO_2	3.42	3.37	0.05	3.85	3.84	0.01
6	OMe	4.47	4.33	0.14	4.36	4.35	0.01

Rate Constant for Elimination Reaction of 2-Z-Phenylethyl p-Y-Arenesulphonates with Potassium t-Butoxide in t-Butyl Alcohol at 40°C. Banger et al. *(9)* reported the rate constants for the elimination reaction of 2-Z-phenylethyl p-Y-arenesulphonate (III) with potassium t-butoxide in t-butyl alcohol. They included twenty-four derivatives in their study: the substituents were H, p-NO_2, p-Me, and p-Br at the Y-position, and the substituents were H, m-OMe, m-Cl, p-Me, p-OMe, and p-Cl at the Z-position.

Fixing each substituent at the Y- or Z-position in turn generates ten subsets as shown in Tables IV and V. From each of these ten subsets, classic QSAR equations 13-22 were derived. Equations 13-16 were derived from the sets fixing the Y-substituents but varying the Z-substituents (Table IV). Equations 17-22 were derived from the sets fixing the Z-substituents but varying the Y-substituents (Table V).



Y = H, p-NO_2, p-Me, p-Br
Z = H, m-OMe, m-Cl, p-Me, p-OMe, p-Cl

(III)

Table IV. Classical QSAR Equations for the Rate Constant of the Elimination Reaction of 2-Z-Phenylethyl p-Y-Arenesulphonates with Potassium t-Butoxide in t-Butyl Alcohol at 40°C Varying Z-Substituents (Z = H, m-OMe, m-Cl, p-Me, p-OMe, p-Cl).

$$\log k = a\,\sigma_Z - b$$

Set	Y	a	b	n	R^2	s	Eq.
1	H	2.50(±0.07)	2.26(±0.02)	6	0.997	0.037	(13)
2	p-Me	2.47(±0.08)	2.40(±0.02)	6	0.995	0.046	(14)
3	p-Br	2.36(±0.06)	1.95(±0.01)	6	0.998	0.030	(15)
4	p-NO_2	2.03(±0.06)	1.35(±0.01)	6	0.996	0.035	(16)

All the correlations are extremely good and highly significant based on their correlation coefficients and standard errors of estimate. The positive coefficients of σ_Z or σ_Y indicate that the electron-withdrawing substituents increase the rate of the elimination reaction. These equations show that the electronic influences of Y- and Z-substituents on the rate of the elimination reaction of substituted arenesulphonates are of different degrees. The average value of the reaction constant ρ for equations 13-16 is 2.3, whereas equations 17-22 have a value equivalent to 1.1; the electronic effects of the same substituent on the rate constant are twice as strong at the Z-position than at the Y-position.

Table V. Classical QSAR Equations for the Rate Constant of the Elimination Reaction of 2-Z-Phenylethyl p-Y-Arenesulphonate with Potassium t-butoxide in t-butyl alcohol at 40°C Varying Y-Substituents (Y = H, p-NO_2, p-Me, p-Br).

$$\log k = a\,\sigma_Y - b$$

Set	Z	a	b	n	R^2	s	Eq.
5	p-OMe	1.24(±0.08)	2.87(±0.03)	4	0.991	0.059	(17)
6	p-Me	1.22(±0.07)	2.64(±0.03)	4	0.994	0.049	(18)
7	p-Cl	0.99(±0.06)	1.66(±0.02)	4	0.994	0.041	(19)
8	H	1.07(±0.04)	2.20(±0.02)	4	0.997	0.028	(20)
9	m-OMe	1.06(±0.04)	1.99(±0.02)	4	0.997	0.030	(21)
10	m-Cl	0.92(±0.04)	1.30(±0.02)	4	0.997	0.027	(22)

In order to investigate the possibility of lateral validation in CoMFA, subsets 5-10 in Table V were first selected. The results of the CoMFA study are summarized in Table VI.

Table VI. CoMFA Equations for the Rate Constant of the Elimination Reaction of 2-Z-Phenylethyl p-Y-Arenesulphonates with Potassium t-butoxide in t-Butyl Alcohol at 40°C Varying Y-Substituents (Y = H, p-NO_2, p-Me, p-Br).[a]

$$\log k = a\ Z1_{H+} - b$$

Set	Z	a	b	n	R^2	s	Eq.
5	p-OMe	0.048(±0.006)	2.614(±0.056)	4	0.969	0.111	(23)
6	p-Me	0.047(±0.008)	2.389(±0.071)	4	0.947	0.142	(24)
7	p-Cl	0.037(±0.008)	1.453(±0.071)	4	0.921	0.142	(25)
8	H	0.041(±0.007)	1.977(±0.066)	4	0.941	0.132	(26)
9	m-OMe	0.040(±0.007)	1.762(±0.068)	4	0.936	0.137	(27)
10	m-Cl	0.031(±0.010)	1.102(±0.095)	4	0.836	0.190	(28)

[a] $Z1_{H+}$ is the first latent variable from an H^+ probe.

The starting coordinates of the molecules were generated using the graphics modeling package for small molecules at Abbott. All geometric variables were optimized with AM1 of AMPAC. The molecules were aligned by superimposing the phenylethyl arenesulphonate moiety. The electrostatic potential energy fields of each molecule were calculated at various lattice points surrounding the molecules using an H^+ probe with the program GRID as described in our previous studies *(7, 8)*. For each molecule, the energies at a total of 2700 grid points were calculated with a 2Å spacing in a lattice of 34 x 28 x 18 (X = -17 to 17, Y = -12 to 16, Z = -11 to 7). All energy values greater than 10 kcal/mol were truncated to 10. Any lattice point for which the standard deviation of the energies was less than 0.05 was discarded. The effects of grid lattice position on the correlation were first investigated by shifting the grid box by -0.5, -1.0, and -1.5Å in each of the directions X, Y, and Z. Based on the results from these different grid box locations, the lattice position that was shifted from the original position by -1.0Å was chosen as the best location of the lattice points.

In general, the quality of the single-component CoMFA models is excellent even though it is not as good as that of the classical QSAR based on their correlation coefficients and standard errors of estimate.

Table VII shows a comparison of the coefficients and the intercepts obtained in the classical QSAR and the CoMFA analyses from the six different sets where the substituent at Z-position is fixed in each case. The average coefficient value of the classical QSAR is 1.1 whereas it is 0.04 in CoMFA 3D-QSAR. Equation 29 shows the relationship between these two sets of coefficients. It is striking to see that there is a high correlation between the coefficient values from the classical QSAR equations and those from the CoMFA as shown in equation 29 as well as Figure 3.

Table VII. Comparison of the Slopes and Intercepts Obtained from CoMFA and Classical QSAR Studies Varying Y-Substituents (Y = H, p-NO_2, p-Me, p-Br).

		Slope		Intercept	
Set	Z	QSAR	CoMFA	QSAR	CoMFA
5	p-OMe	1.2411	0.0482	-2.8746	-2.6056
6	p-Me	1.2170	0.0466	-2.6449	-2.3893
7	p-Cl	0.9946	0.0374	-1.6621	-1.4532
8	H	1.0703	0.0409	-2.2016	-1.9768
9	m-OMe	1.0627	0.0402	-1.9851	-1.7619
10	m-Cl	0.9244	0.0311	-1.2966	-1.1025

$$\text{Slope (QSAR)} = 19.537(\pm 1.709)\ \text{Slope (CoMFA)} + 0.289(\pm 0.070) \qquad (29)$$
$$n = 6,\ r^2 = 0.970,\ s = 0.024,\ F = 130.7,\ p = 0.0003$$

The results again indicate that the coefficients of CoMFA PLS correlations provide information that is comparable to that of the classical QSAR correlations and suggest their possible use in the lateral validation in CoMFA.

A high correlation (equation 30) also exists between the intercepts obtained from the two different approaches as illustrated in Figure 4. Similar results were obtained previously from the comparison of the classical QSAR and CoMFA correlations with the data of log MKC values of bromoalkylcarboxylic acids *(10)*.

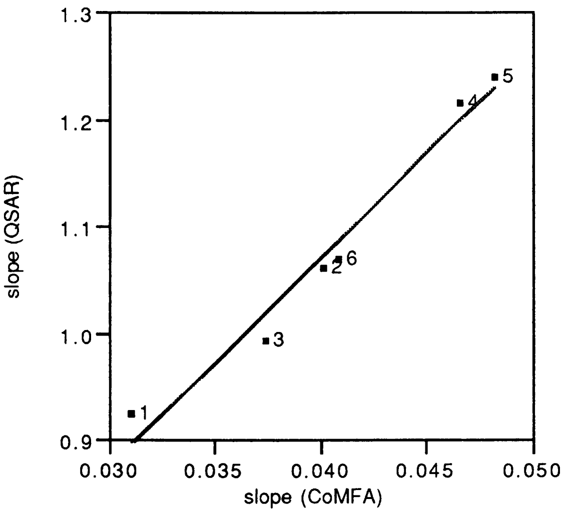


Figure 3. Plot between the reaction constants ρ (slope) obtained in classical QSAR and those in CoMFA.

$$\text{Intercept (QSAR)} = 1.046(\pm 0.005)\ \text{Intercept (CoMFA)} - 0.141(\pm 0.009) \qquad (30)$$
$$n = 6,\ r^2 = 0.9999,\ s = 0.006,\ F = 47674,\ p = 0.000$$

During the course of this CoMFA study, a wrong set of data was once unintentionally chosen. Nevertheless, it turned out that this provided additional information. Equations 31 and 32 are the correlations obtained from the classical QSAR and CoMFA from four Z-substituted derivatives (Y = H, Z = H, m-Cl, m-OMe, and p-Cl). In this case, the Y-substituent is fixed instead of the Z-substituent. The coefficient value from the classical QSAR (equation 31) is 2.4 which is similar to that of equation 13, and about twice as large as those of equations 17-22 as before. The coefficient value from CoMFA (equation 32) is also larger than those of equations 23-28.

$$\log k = 2.44(\pm 0.18)\ \sigma_Z - 2.25(\pm 0.04) \qquad (31)$$
$$n = 4,\ R^2 = 0.989,\ s = 0.049$$

$$\log k = 0.110(\pm 0.047)\ Z1_H^{+} - 1.807(\pm 0.122) \qquad (32)$$
$$n = 4,\ R^2 = 0.735,\ s = 0.244$$

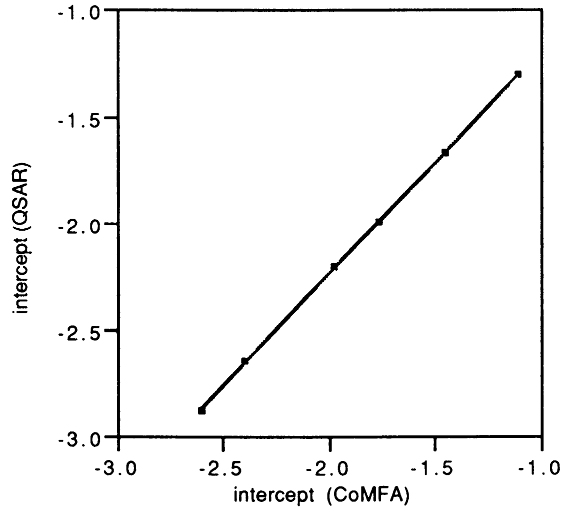


Figure 4. Plot between the intercepts obtained in classical QSAR and those in CoMFA of the sets described in Table VII.

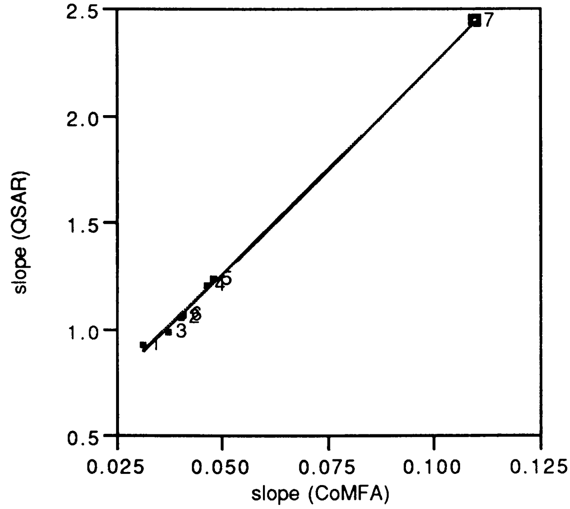


Figure 5. Plot between the intercepts obtained in classical QSAR and those in CoMFA of the sets described in Table VII including one point (Labeled 7) described in Equations 31 and 32.

This new data point was added to those used in equation 29, and a correlation was derived (equation 33). Equation 33 and Figure 5 show the relationship between the coefficients of QSAR and CoMFA correlations from all seven data points. They show that the new point labeled 7 did not alter the relationship described in equation 29; the coefficients and the intercepts of equations 29 and 33 are essentially identical.

$$\text{Slope (QSAR)} = 19.561(\pm 0.325)\ \text{Slope (CoMFA)} + 0.288(\pm 0.018) \quad (33)$$
$$n = 7,\ r^2 = 0.999,\ s = 0.021,\ F = 3627.5,\ p = 0.0000$$

The results described in equation 33 provide additional support for the potential use of CoMFA PLS regression coefficients in lateral comparisons.

Miscellaneous Aspects of Lateral Comparison

Application in Combining Multiple Equations. Other applications for the coefficients in PLS equations include their use in combining multiple equations. From the log ($1/LD_{100}$) and log ($1/LD_{50}$) data of a set of 11 and/or 6 benzyldimethylalkylammonium derivatives that was used to derive a parabolic equation in the classical QSAR, a two-component model was derived from CoMFA *(11)* (equations 34-40). LD_{100} and LD_{50} in these equations are the measured highest molar tolerated concentration (100%) of the male mice after continuous infusion at a constant speed and the molar acute intravenous toxic concentration in 50% of the male mice after bolus injection, respectively. The figures in brackets show the percentage of variance in the biological activity with respect to the first, second, and third variables in the corresponding equations. Equations 34-36 are from the classical QSAR, and equations 37-40 are from CoMFA. The coefficients of log P and $(\log P)^2$ in equations 34 and 35 are essentially identical but with only different intercepts. Thus, these equations can be combined into one using an indicator variable I_{LD50} as shown in equation 36. Likewise, the coefficients of $Z1_{H2O}$ and $Z2_{H2O}$ in equations 37 and 38 are similar once again with only their intercepts differing. $Z1_{H2O}$ and $Z2_{H2O}$ in these equations are the first and second latent variables from an H_2O probe, respectively. Thus, as in the classical QSAR, equations 37 and 38 can be combined into one as shown in equations 39 and 40.

$$\log 1/LD_{100} = 0.939 \log P - 0.107(\log P)^2 + 0.200 \quad (34)$$
$$n = 11,\ r^2 = 0.864,\ s = 0.148,\ [51\%, 35\%]$$

$$\log 1/LD_{50} = 0.945 \log P - 0.105(\log P)^2 + 0.563 \quad (35)$$
$$n = 6,\ r^2 = 0.975,\ s = 0.074,\ [50\%, 45\%]$$

$$\log 1/LD = 0.941 \log P - 0.107(\log P)^2 + 0.453 I_{LD50} + 0.166 \quad (36)$$
$$n = 17,\ r^2 = 0.917,\ s = 0.128,\ [36\%, 24\%, 31\%]$$

$$\log 1/LD_{100} = 0.038 Z1_{H2O} + 0.038 Z2_{H2O} + 1.940 \quad (37)$$
$$n = 11,\ r^2 = 0.944,\ s = 0.095,\ r^2_{cv} = 0.892,\ s_{cv} = 0.170,\ [68\%, 26\%]$$

$$\log 1/LD_{50} = 0.035 Z1_{H2O} + 0.036 Z2_{H2O} + 2.348 \quad (38)$$
$$n = 6,\ r^2 = 0.981,\ s = 0.064,\ r^2_{cv} = 0.708,\ s_{cv} = 0.274,\ [67\%, 31\%]$$

$$\log 1/LD = 0.037 Z1_{H2O} + 0.037 Z2_{H2O} + 0.408 I_{LD50} + 1.940 \quad (39)$$
$$n = 17,\ r^2 = 0.966,\ s = 0.082,\ r^2_{cv} = 0.942,\ s_{cv} = 0.092,\ [51\%, 21\%, 25\%]$$

$$\log 1/LD = 0.036 Z1_{H2O} + 0.039 Z2_{H2O} + 0.429 I_{LD50} + 1.931 \quad (40)$$
$$n = 17,\ r^2 = 0.965,\ s = 0.083,\ r^2_{cv} = 0.940,\ s_{cv} = 0.094,\ [49\%, 20\%, 27\%]$$

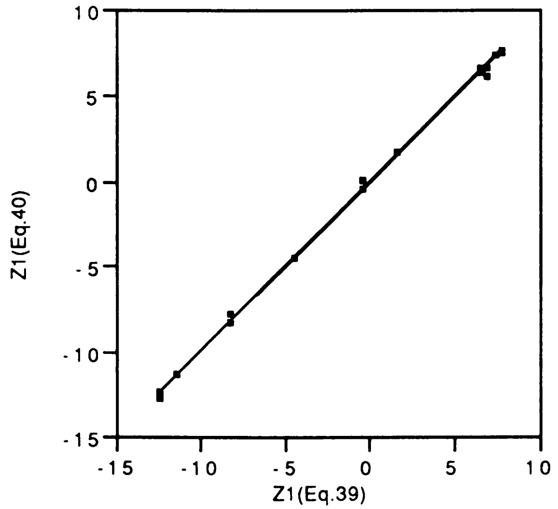


Figure 6. Plot between Z1 from equation 39 (n=11 and n=6) and Z1 from equation 40 (n=17).

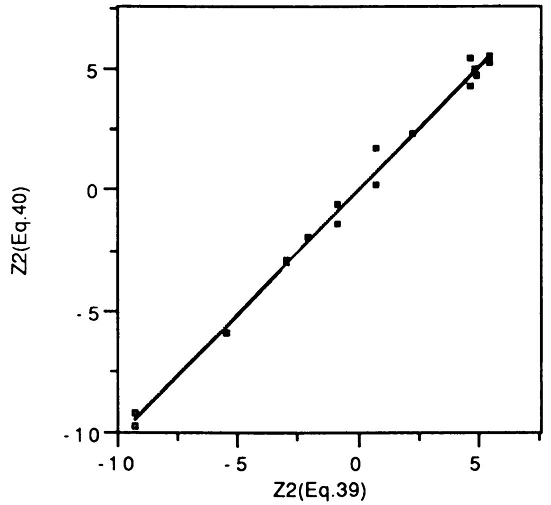


Figure 7. Plot between Z2 from equation 39 (n=11 and n=6) and Z2 from equation 40 (n=17).

The values of latent variables in equation 39 are the same as those used in equations 37 and 38, and the values of latent variables in equation 40 are derived from all seventeen compounds in PLS. However, they are essentially identical as shown in equations 41 (Z1) and 42 (Z2) and Figures 6 (Z1) and 7 (Z2).

Table VIII. Comparison of the Coefficients and Intercepts of the CoMFA Results from Different Grid Positions for the Rate Constant of the Elimination Reaction of 2-Z-Phenylethyl p-Y-Arenesulphonates with Potassium t-Butoxide in t-Butyl Alcohol at 40°C Varying Y-Substituents (Y = H, p-NO_2, p-Me, p-Br).

Set	Z	Grid Shift	Coefficient	Intercept	R^2	s	s_{cv}
		0.0	0.0319 (±0.0100)	-1.1025 (±0.0950)	0.8360	0.1901	0.6380
		-0.5	0.0319 (±0.0100)	-1.1025 (±0.0950)	0.8362	0.1900	0.6420
1	m-Cl	-1.0	0.0311 (±0.0097)	-1.1025 (±0.0949)	0.8365	0.1898	0.6283
		-1.5	0.0308 (±0.0097)	-1.1025 (±0.0952)	0.8355	0.1904	0.6327
		0.0	0.0408 (±0.0077)	-1.7619 (±0.0695)	0.9337	0.1389	0.2393
		-0.5	0.0401 (±0.0077)	-1.7619 (±0.0703)	0.9320	0.1407	0.2302
2	m-OMe	-1.0	0.0402 (±0.0075)	-1.7619 (±0.0684)	0.9356	0.1369	0.2267
		-1.5	0.0395 (±0.0075)	-1.7619 (±0.0696)	0.9334	0.1392	0.2280
		0.0	0.0388 (±0.0077)	-1.4532 (±0.0687)	0.9261	0.1375	0.4601
		-0.5	0.0382 (±0.0079)	-1.4532 (±0.0708)	0.9216	0.1416	0.4815
3	p-Cl	-1.0	0.0374 (±0.0077)	-1.4532 (±0.0711)	0.9210	0.1421	0.5131
		-1.5	0.0374 (±0.0077)	-1.4532 (±0.0704)	0.9225	0.1408	0.5281
		0.0	0.0478 (±0.0078)	-2.3893 (±0.0697)	0.9492	0.1394	0.4370
		-0.5	0.0474 (±0.0080)	-2.3893 (±0.0722)	0.9455	0.1444	0.4403
4	p-Me	-1.0	0.0466 (±0.0078)	-2.3893 (±0.0710)	0.9473	0.1420	0.4770
		-1.5	0.0463 (±0.0077)	-2.3893 (±0.0711)	0.9472	0.1422	0.4765
		0.0	0.0494 (±0.0063)	-2.6140 (±0.0563)	0.9683	0.1126	0.2896
		-0.5	0.0498 (±0.0065)	-2.6140 (±0.0570)	0.9674	0.1140	0.2713
5	p-OMe	-1.0	0.0482 (±0.0061)	-2.6140 (±0.0556)	0.9690	0.1112	0.2865
		-1.5	0.0489 (±0.0062)	-2.6140 (±0.0557)	0.9689	0.1114	0.2761
		0.0	0.0418 (±0.0075)	-1.9768 (±0.0664)	0.9401	0.1329	0.3729
		-0.5	0.0418 (±0.0075)	-1.9768 (±0.0664)	0.9402	0.1329	0.3477
6	H	-1.0	0.0409 (±0.0073)	-1.9768 (±0.0662)	0.9406	0.1324	0.3417
		-1.5	0.0407 (±0.0073)	-1.9768 (±0.0669)	0.9393	0.1338	0.3513

$$Z1_{H2O}(\text{eq.}40) = 1.011(\pm 0.009)\ Z1_{H2O}(\text{eq.}39) + 0.000(\pm 0.007) \qquad (41)$$
$$n = 17,\ r^2 = 0.999,\ s = 0.278,\ F = 12800.6,\ p = 0.0001$$

$$Z2_{H2O}(\text{eq.}40) = 0.972(\pm 0.021)\ Z2_{H2O}(\text{eq.}39) - 0.000(\pm 0.103) \qquad (42)$$
$$n = 17,\ r^2 = 0.993,\ s = 0.423,\ F = 2105.6,\ p = 0.0001$$

Influence of the Position of Lattice Points on the Coefficients. Even though it would be possible to use the coefficients in PLS regression in lateral validation or lateral comparison, it would be impractical if the coefficients varied depending on the various CoMFA conditions such as the position of lattice points. Table VIII shows a comparison of the coefficients and intercepts of the CoMFA results from different grid positions for the rate constant of the elimination reaction of 2-Z-phenylethyl p-Y-arenesulphonates varying Y-substituents. Table IX shows a comparison of the coefficients and intercepts from different grid positions for the log MIC (minimum inhibitory concentration) or log MKC (minimum killing concentration) of 6 to 12 benzyldimethylalkylammonium chlorides against various microbial strains *(6)*. These results show that the shifts in the grid box have little effect on the coefficients and slopes; the comparison of coefficients in PLS regression equations is not likely to be significantly affected by the change of the grid box.

Table IX. Comparison of the Coefficients and Intercepts of the CoMFA Results from Different Grid Positions for the Log MIC (Minimum Inhibitory Concentration) or Log MKC (Minimum Killing Concentration) of 6 to 12 Benzyldimethylalkylammonium Chlorides against Various Microbial Strains (See reference 6).

No	Grid Shift	Coefficients		Intercept
		$Z1_{H2O}$	$Z2_{H2O}$	
		P. aeruginosa. Log MIC		
1A-a	1.0	0.066(±0.008)	0.019(±0.006)	3.068(±0.040)
1A-b	0.5	0.053(±0.005)	0.015(±0.004)	3.068(±0.031)
1A	0.0	0.059(±0.006)	0.012(±0.004)	3.068(±0.034)
1A-c	-0.5	0.055(±0.007)	0.015(±0.005)	3.068(±0.040)
		S. typhosa, Log MIC		
2A-a	1.0	0.055(±0.006)	0.048(±0.009)	4.300(±0.042)
2A-b	0.5	0.048(±0.002)	0.036(±0.004)	4.300(±0.021)
2A	0.0	0.049(±0.003)	0.040(±0.004)	4.300(±0.022)
2A-c	-0.5	0.050(±0.002)	0.040(±0.004)	4.300(±0.019)
		P. vulgaris Log MIC		
3A-a	1.0	0.053(±0.007)	0.053(±0.010)	3.340(±0.047)
3A-b	0.5	0.049(±0.003)	0.038(±0.004)	3.340(±0.027)
3A	0.0	0.048(±0.004)	0.043(±0.005)	3.340(±0.031)
3A-c	-0.5	0.049(±0.004)	0.043(±0.005)	3.340(±0.026)
		P. vulgaris Log MKC		
4A-a	1.0	0.104(±0.014)	0.022(±0.008)	3.223(±0.060)
4A-b	0.5	0.089(±0.005)	0.011(±0.003)	3.223(±0.027)
4A	0.0	0.097(±0.005)	0.008(±0.003)	3.223(±0.029)
4A-c	-0.5	0.094(±0.006)	0.011(±0.003)	3.223(±0.031)
		S. aureus Log MIC		
5A-a	1.0	0.071(±0.005)	0.081(±0.008)	5.118(±0.038)
5A-b	0.5	0.063(±0.004)	0.059(±0.005)	5.118(±0.033)
5A	0.0	0.064(±0.003)	0.063(±0.004)	5.118(±0.022)
5A-c	-0.5	0.064(±0.004)	0.066(±0.006)	5.118(±0.035)
		Cl. welchii Log MIC		
6A-a	1.0	0.082(±0.005)	0.082(±0.009)	4.410(±0.038)
6A-b	0.5	0.071(±0.003)	0.059(±0.005)	4.410(±0.027)
6A	0.0	0.071(±0.003)	0.066(±0.005)	4.410(±0.025)
6A-c	-0.5	0.074(±0.003)	0.065(±0.006)	4.410(±0.032)
		P. aeruginosa Log MKC		
7A-a	1.0	0.081(±0.006)	0.077(±0.008)	2.967(±0.042)
7A-b	0.5	0.073(±0.004)	0.055(±0.004)	2.967(±0.029)
7A	0.0	0.077(±0.003)	0.057(±0.003)	2.967(±0.022)
7A-c	-0.5	0.076(±0.002)	0.060(±0.003)	2.967(±0.018)
		Cl. welchii Log MKC		
8A-a	1.0	0.087(±0.004)	0.088(±0.008)	4.384(±0.036)
8A-b	0.5	0.076(±0.003)	0.063(±0.005)	4.384(±0.029)
8A	0.0	0.076(±0.002)	0.071(±0.004)	4.384(±0.022)
8A-c	-0.5	0.078(±0.003)	0.070(±0.006)	4.384(±0.032)

Table IX. *Continued*

		Red cell sheep	Log C(H50)	
9A-a	1.0	0.122(±0.009)	0.117(±0.018)	3.903(±0.077)
9A-b	0.5	0.106(±0.011)	0.084(±0.021)	3.903(±0.110)
9A	0.0	0.109(±0.007)	0.081(±0.012)	3.903(±0.067)
9A-c	-0.5	0.109(±0.009)	0.090(±0.017)	3.903(±0.087)
		C. albicans	Log MKC	
10A-a	1.0	0.085(±0.009)	0.094(±0.013)	3.979(±0.062)
10A-b	0.5	0.081(±0.006)	0.063(±0.007)	3.979(±0.043)
10A	0.0	0.082(±0.005)	0.069(±0.006)	3.979(±0.038)
10A-c	-0.5	0.083(±0.005)	0.068(±0.006)	3.979(±0.040)
		Red cell sheep	Log C(H50)	
11A-a	1.0	0.116(±0.005)	0.076(±0.011)	3.707(±0.046)
11A-b	0.5	0.100(±0.006)	0.052(±0.012)	3.707(±0.062)
11A	0.0	0.102(±0.003)	0.053(±0.006)	3.707(±0.033)
11A-c	-0.5	0.103(±0.005)	0.057(±0.009)	3.707(±0.046)

Conclusions

The results obtained from the comparative studies of the dissociation constants of benzoic acids and phenylacetic acids and the rate constants for the elimination reaction of 2-Z-phenylethyl p-Y-arenesulphonates indicate that the coefficients of the PLS regression equations in CoMFA contain useful information. Specifically, a potential application for the coefficients can be found in the lateral validation or lateral comparison of single-component models. However, even though the present results are very encouraging, a comparison of the coefficients in normal CoMFA studies, where the optimum number of components (latent variables) is greater than one, may not be so simple. It can become complicated because the optimum number of components varies depending on the constitution of compounds included in the analysis. Further investigation is required to generalize the applicability of the coefficients in PLS regression equations in lateral validation.

References

1. Kim, K. H. In: Kubinyi H, ed. 3D QSAR in Drug Design. Theory Methods and Applications. Leiden: ESCOM, 1993: 619.
2. Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147.
3. Hansch, C.; Klein T. *Acc. Chem. Res.* **1986**, *19*, 392.
4. Hansch, C.; Bjorkroth, J. P.; Leo, A. *J. Pharm. Sci.* **1987**, *76*, 663.
5. Shorter, J. Correlation Analysis in Organic Chemistry: an introduction to linear free-energy relationships.Oxford: Clarendon, 1973, pp. 1-2.
6. Kim, K. H. *J. Comput.-Aid. Mol. Design* **1993**, *7*, 71.
7. Kim, K. H.; Martin, Y.C. *J. Med. Chem.* **1991**, *34*, 2056.
8. Kim, K. H.; Martin, Y. C. *J. Org. Chem.* **1991**, 56, 2723.
9. Banger, J; Cockerill, A.F; Davies, G. L. O. *J. Chem. Soc. B.* **1971**, 498.
10. Kim, K. H. *Quant. Struct.-Act. Relat.* **1992**, *11*, 309.
11. Kim, K. H. *Med. Chem. Res.* **1993**, *3*, 257.

RECEIVED April 26, 1995

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Chapter 24

Distance Comparisons: A New Strategy for Examining Three-Dimensional Structure–Activity Relationships

Yvonne Connolly Martin

Computer Assisted Molecular Design Department, Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, IL 60064

DISCO is a pharmacophore mapping method that quickly generates all pharmacophore models consistent with the data. This provides an overview of the pharmacophore maps just as all possible regressions provides an overview of the QSAR of the dataset. In many cases more than one pharmacophore map is consistent with the data: additional experimental information must be supplied to resolve the issue. Frequently, for some of the compounds in the dataset DISCO identifies more than one possible bioactive conformation. These properties make DISCO a complement to 3D QSAR methods.

The objective of computer-assisted molecular design is to use the computer to discover novel, patentable compounds. Frequently, the only information available for this analysis is the structure-activity relationships within a set of molecules. From such information one attempts to understand the physical properties and molecular shape associated with the biological property. In this report I will describe some new methods developed in our group to solve some of the problems presented to a computational chemist working in the pharmaceutical and agrochemical industry.

These new methods complement, and do not replace, the more familiar methods of drug design aided by Hansch-Fujita [Hansch and Fujita, 1964] and 3D QSAR [Cramer III, et al., 1988], molecular graphics [Langridge, et al., 1981], manual pharmacophore mapping [Marshall, et al., 1979], or macromolecular structure determination [Goodford, 1984]. Indeed, by using different tools one gains flexibility in the concepts examined and, if the results from different methods are consistent with each other, confidence that the hypotheses are reasonable.

Hansch-Fujita QSAR: Strengths and Limitations.

Traditional QSAR involves the statistical analysis of the relationship between molecular descriptors, usually physical properties, and biological potency [Hansch and Fujita, 1964, Martin, 1978]. It is a powerful technique for the computer analysis of structure-activity relationships. If descriptors for the molecules are available, a QSAR analysis can often be accomplished rather quickly. Furthermore, QSAR has

0097–6156/95/0606–0318$12.00/0

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

been proven to correctly forecast the potency of molecules before their synthesis [Martin, 1981, Fujita, 1984]. Indeed, several marketed compounds were discovered with QSAR [Boyd, 1990]. Lastly, the QSAR viewpoint also provides strategies for the rational design of the series to follow up a lead [Hansch, et al., 1973].

A very attractive feature of QSAR is that a careful analysis involves the comparison of many hypotheses in the form of different equations. Sometimes the analysis reveals that one equation is superior to the others considered. In other cases, two or more equations involving different physical properties result from the analysis. The problem is then to distinguish which, if either, of the equations is correct.

Since the properties used to describe molecules for a QSAR are based on experimental measurements, it frequently happens that one cannot estimate the physical properties of interest for novel structures. Such molecules cannot therefore be used in a QSAR analysis. Additionally, it can be difficult to merge separate series of very different parent structure for a common QSAR analysis. Thus although QSAR is a truly powerful method, its reliance on physicochemical descriptors may prevent its use in cases of interest.

3D QSAR: Strengths and Limitations.

3D QSAR methods were developed as an alternative to Hansch-Fujita QSAR to describe molecules more "realistically", that is with properties of molecules calculated from their three-dimensional structures generated by molecular modeling [Kubinyi, 1993]. For most 3D QSAR methods, one first selects a bioactive conformation for each molecule and aligns them in a common 3D space. In the most popular method, CoMFA (*Co*mparative *M*olecular *F*ield *A*nalysis) [Cramer III, et al., 1988], the molecular descriptors are steric, electrostatic, and sometimes hydrogen-bonding fields calculated at the intersections of a 3D lattice. The statistical relationships between these fields and biological potency is analyzed with the PLS (*P*artial *L*east *S*quares) method. As with traditional QSAR, the statistical basis of CoMFA provides a tool to identify and compare how different hypotheses fit the same data.

For cases for which it has been examined, traditional QSAR and CoMFA produce statistically equivalent results. Additionally, fields generated with the program Grid reproduce traditional QSAR steric, electrostatic, and hydrophobic properties of molecules [Kim, 1993]. Hence, CoMFA has at least partially solved the problem of generating descriptors for unusual molecules. Furthermore, it correctly forecasts the potency of molecules not included in the derivation of a relationship [Martin, et al., in press].

However, to apply CoMFA one must solve another problem, namely how to align the molecules in 3D. One must decide what points in molecule A correspond to which points in molecule B, etc. A part of this "alignment problem" is the selection of the bioactive conformation of each molecule. Thus, a critical preliminary to 3D QSAR is to map the pharmacophore, that is to propose the groups required for biological activity and the distances between them. Generally, but not always, a 3D QSAR will be applied to only those molecules that match the pharmacophore since the absence of any pharmacophore feature is hypothesized to be sufficient to render a molecule inactive.

Of course, if all the structures in the dataset contain a common backbone, one may choose to superimpose the molecules based on this backbone and to use any arbitrary conformation for the analysis. However, such a superposition may require that the 3D structures of some molecules are in a high energy conformation. Furthermore, extrapolations to novel structural backbones are not necessarily reliable.

This report will first provide a summary of our automated pharmacophore mapping program DISCO, *Dis*tance *Co*mparisons, then illustrate how it complements 3D QSAR [Martin, et al., 1993].

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Overview of the DISCO Strategy for Pharmacophore Mapping.

Step 1. Identify the Set of Active Compounds on Which to Base the Pharmacophore. Following the precedent of Marshall and colleagues, DISCO searches for a pharmacophore map in only the potent molecules [Marshall, et al., 1979]. The hypothesis behind ignoring the inactive compounds in the first stages of a structure-activity analysis is that many factors can render a compound inactive. Some related molecules are inactive because they do not contain the correct pharmacophore hydrogen-bonding or charge centers; others cannot achieve the correct distances between the required features; others are inactive because, in conformers that match the pharmacophore distances, they occupy space that is required by the target biomolecule; and still others are inactive because the molecule has the wrong physical properties--such as incorrect charges on key atoms or the wrong overall octanol-water log P. As an example, Figure 1 shows examples of 2-aminotetralins that bind (a) or do not bind (b-c) to the D2 dopamine receptor.

Our CoMFA model for D1 dopaminergic agonists [Martin, et al., 1993], Figure 2, underscores the importance of the shape of molecules on their receptor affinity. For this model, only steric energies contribute to affinity. Significantly, this model has useful precision in forecasting potency of molecules before their synthesis. Table I summarizes the forecast and observed affinities for molecules modeled before their synthesis. First note that although we modeled 201 compounds, only 19 were synthesized. Half of these were forecast to have significant binding affinity, the other half were made either to answer important structure-activity questions or because they were by-products of other syntheses. Five of the molecules forecast to bind strongly indeed do so. In contrast, none of the molecules forecast not to bind were observed to bind.

Table I. The Forecast D1 Receptor Affinity of Molecules Proposed for Synthesis.

CoMFA Forecast pKi	Number Compounds		Observed pKi		
	Modeled	Synthesized	<6	6–7	>7
< 6.5	146	9	6	3	0
> 6.5	55	10	1	4	5

Step 2. Generate All Low-energy Conformers to Consider as Potentially the Bioactive One. Both because it is usually not possible to accurately calculate the relative energies of different conformers and because the bound conformation of a ligand will reflect the total energy of the system and so not be the global minimum energy of the ligand alone, it is important to base a pharmacophore mapping strategy on consideration of all accessible conformations of the active molecules. Even though it is attractive to model the 3D structure of one molecule from the selected bioactive conformer of another, the danger in this strategy is that the selected conformer may be a high-energy one for the analogue. We typically generate the conformers with distance geometry [Blaney, et al., 1990] and minimize the structures with MMP2 [Allinger, 1987] or AM1 [Dewar, et al., 1985]. Our computer program Reject identifies the lowest energy conformer, a local minimum conformer, of a group of conformers that differ by no more than some tolerance, typically 0.3Å, in all distances between heavy atoms. For example, Figure 3 shows a set of nicotinic agonists [Sheridan, et al., 1986] and the number of low-energy conformations found for even these relatively rigid compounds.

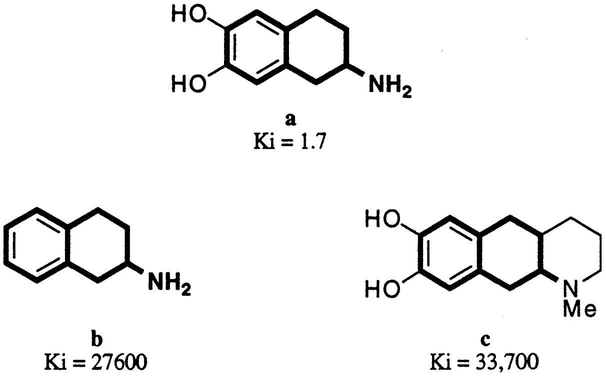


Figure 1. The structures of aminotetralins tested for binding to the D2 receptor.

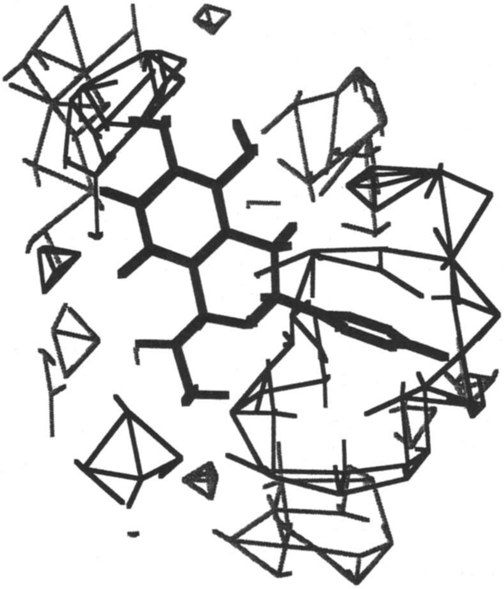


Figure 2. The steric CoMFA model for D1 affinity (n = 61; fitted statistics: R^2 = 0.86, root-mean square error = 0.43; cross-validated statistics: R^2 = 0.51, root-mean square error = 0.79). The positive contours, regions where increasing bulk increase affinity, are shown as heavy lines and the negative contours are shown as lighter ones.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Step 3. Label Charged, Hydrogen-bonding, and Hydrophobic Points that May Form Part of the Superposition Rule. Whereas early pharmacophore maps were based on superpositions of atoms, experience has shown that better results are obtained when one also includes projections from ligand atoms to site points, pseudoatoms, that indicate the idealized location of the complementary protein or nucleic acid atoms that would interact with the ligand atoms.

In DISCO the locations of these site points are based on observations of the geometry of hydrogen bonds and salt bridges in small-molecule crystals. They also reflect the location of minima and maxima in the electrostatic potential surface of a molecule.

In principle the site points are located at extensions of the heavy-atom--lone-pair line. However, a simple modification allows one to consider all rotamers from one 3D structure of the heavy atoms, illustrated in Figure 4. Since basic nitrogens protonate before forming a salt bridge, there are three locations for potential anionic sites associated with any conformer of a primary amine and two with a secondary amine. In a similar vein, for any one conformer of an alcohol, there are three locations of potential hydrogen-bond acceptor points and three for potential hydrogen-bond donor points.

Potential site points are considered in a particular conformation only if they are outside the van der Waals surface and thus accessible to intermolecular contact. The points that DISCO considers for superposition are under user control--one can exclude certain default points or add both new ligand points (halogens, for example) and new site points (to idealized locations of metal atoms, for example).

Step 4. Find the Superpositions and Conformers that Match in 3D. The heart of DISCO is the Bron-Kerbosh clique-detection algorithm [Bron and Kerbosch, 1973]. It was shown to be the fastest of several tested to find the maximum common 3D substructure in two 3D structures [Brint and Willett, 1987]. In the case of pharmacophore mapping, the common 3D substructure is the pharmacophore map consisting of ligand and site points.

A simple maximum common 3D substructure is not sufficient since a pharmacophore map is built from more than two structures and more than one conformer of a flexible compound may match a rigid one. Hence, DISCO follows a sequence of comparisons to produce the set of unique solutions. The compound with the fewest conformations is used as the reference. Each conformer of every other molecule is compared to each conformer of it.

The user sets the criteria for a match between two 3D structures. The first is the points that must or may be used in the superposition. Next is the tolerance within which two distances will be considered identical. Table II shows an empirical experiment to demonstrate that, depending on the number of points considered, the tolerance is roughly double the root-mean-square difference for superposition. The reference "molecule" had one point at the center and six points at the intersections of a cube 8Å on a side. For the first test "molecule", one of the points was moved 1.0 Å along an axis. If one superimposes the center and two intersections of the first with the center, the moved point, and one remaining point of the second, the resulting root-mean-square superposition is 0.46Å. For the second and subsequent test "molecules", the next point was moved as much as possible while still maintaining all inter-"molecule" distances $\leq$1.0Å. The columns in the table describe the effect of adding more points to the superposition.

The user also specifies the increment to add to the tolerance if no pharmacophore map is found as well as the upper bound at which searching is discontinued. Other optional criteria are a relative energy cut-off for considering a conformation, that a certain number of points of no specified type must be included, that some specified compounds (inactive analogues, for example) need not match the pharmacophore map, or that the map need not include every compound.

DISCO reports a pharmacophore map if at least one conformer of the required

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

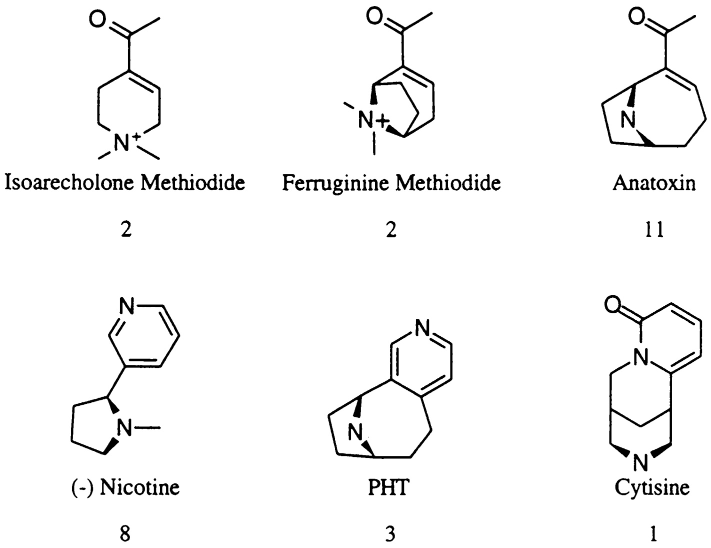

Isoarecholone Methiodide
2
Ferruginine Methiodide
2
Anatoxin
11
(-) Nicotine
8
PHT
3
Cytisine
1

Figure 3. The structures and number of unique low-energy conformations of some nicotinic agonists.

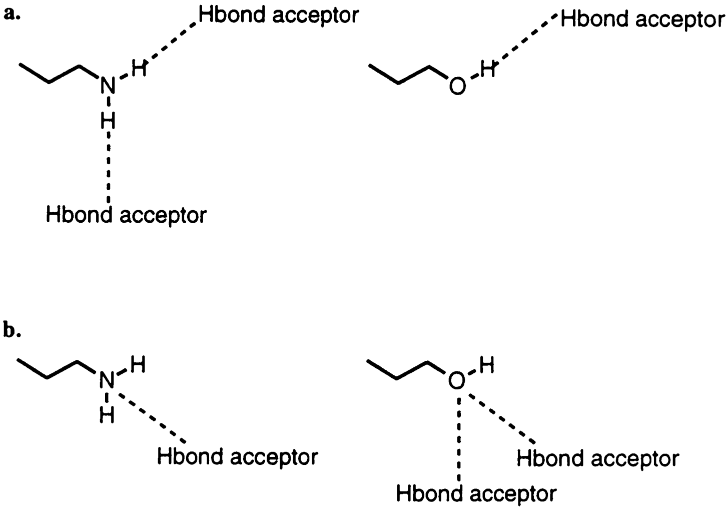


Figure 4. Location of site hydrogen bond acceptor points of a primary amine and an alcohol. In panel *a* the locations of the site points are calculated from the location of the hydrogen-bonding proton. In panel *b* the locations of the site points are calculated from the location of the lone pairs on the basis that the amine would protonate at this position and the alcohol could rotate about the C–OH bond to place a proton at this position. For the alcohol site, hydrogen-bond donor points are also placed at the same three positions.

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Table II. The Relationship Between the Tolerance for Clique-detection and the Root-mean-square Values for Superposition.

Number of Atoms Moved 1.0 Å	Number of Points Superimposed 3	4	5	6
1	0.46	0.43	0.40	0.37
2	0.50	0.47	0.43	0.40
3	0.54*	0.48	0.44	0.40
4	0.47	0.50	0.45	0.41
5	0.47	0.50	0.53	0.49
6	0.47	0.50	0.49	0.50

* A 2Å shift results in an 0.98 RMS.

molecules match the reference according to the criteria set. If no matches are found, then the tolerance is incremented and the search is repeated.

Step 5. Evaluate the DISCO Results. The "best" pharmacophore map is that which overlaps the most points of all active molecules, doesn't use high-energy conformers, explains why other compounds are inactive, and leads to a predictive 3D QSAR.

Note that the process of examining the possible pharmacophore maps is analogous to examining the regression equations for QSAR. Both the objective qualities of the respective models and the subjective experience of the investigator must be considered. As with regression analysis, a systematic analysis of possible pharmacophore maps may find no solution. Then either an additional property, another type of point or more conformers, must be added or some molecules dropped from the model. Again by analogy to regression analysis, a systematic analysis of possible pharmacophore maps may find two solutions that are not distinguishable with the data presented. A follow-up analysis of all the compounds tested may increase the likelihood of one of the solutions. However, it may be necessary to test additional compounds to distinguish between the hypotheses.

From observations of the 3D structures of ligand-macromolecule complexes, we consider a tolerance of 2Å to be reasonable. We put more weight on the superposition of site points than of ligand points, although the directionality of hydrogen-bonds must be considered. The energy cut-off is more complicated in that for many molecules one may not be convinced that the calculated relative energies are valid. This is especially true if some conformers of a molecule have an intramolecular hydrogen-bond that both lowers the energy of the isolated molecule but must be overcome in forming the ligand-macromolecule complex.

For example, consider the pharmacophore maps found by running DISCO using different criteria for the dopamine D1 agonists shown in Figure 5. Table III shows that the default settings produced a map with 0.8 Å tolerance. It superimposes four points in the D1 agonists--the basic nitrogen, the meta-OH group, and site points calculated from each. By requiring that the phenolic phenyl ring is also superimposed, DISCO finds one solution at a tolerance of 1.0 Å. This solution also contains only very low-energy conformers. To include the pendant phenyl ring a tolerance of 1.8 Å is needed. This solution also includes the para-oxygen atom. However, no sensible solution is found that also includes the site point from the para-oxygen.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Table III. DISCO Pharmacophore Maps for Diverse Dopamine D1 Agonists.

Tolerance Δ Å	Largest Minimum. Δ E kcal/mole	Points Included Ligand	Points Included Site
0.8	0.85	N, m-O	N, m-O
1.0	0.02	N, m-O, @	N, m-O
1.8	0.02	N, m & p-O, @, #	N, m-O
>2.0	>3.0	N, m & p-O, @, #	N, m & p-O

@ centroid of the phenolic ring.
# centroid of the unsubstitued phenyl ring.

A requirement of any method that will be used to explore alternative hypotheses is that it supply answers promptly. Hence, it is of interest that each test in Table III required one minute on a Vax 9000 to compare seven molecules, 5880 combinations of conformations per set of points, and 6-16 points per conformation. All the data in Table III was generated (once the conformers had been generated and labeled) in four minutes. Clearly, such quick examination of hypotheses is an aid to molecular modeling.

Example: Comparison of Nicotinic Agonists with DISCO[Holladay, et al., 1994].

An example of the use of DISCO is the attempted superposition of two nicotinic agonists, nicotine and its methyl-isoxazole analogue, ABT-418 [Garvey, et al., 1994]. Two discretely different maps were found, Figure 6. Note that this would be equivalent in regression analysis to finding two equations that explain the data. As in regression analysis, the solution is found by adding another molecule to the dataset. In this case the nicotinic agonist PHT fits only one of the hypotheses, Figure 7.

The utility of DISCO is shown when acetyl choline, the natural ligand for the nicotinic receptor, is added to the model. Several conformers of acetyl choline match the model, Figure 8. Which (one) is appropriate to use for 3D QSAR?

A similar circumstance arises with the N-des-methyl analogue of ABT-418. Removing the methyl group results in much freer rotation of the isoxazole group. The same is not true of the corresponding pyridyl analogue, N-des-methyl nicotine. In the case of the former compound, several low-energy conformers match the pharmacophore, Figure 9. For a 3D QSAR should one choose a low-energy conformer that occupies new volume or the conformer 1.85 kcal/mole higher that overlaps the selected conformation of ABT-418, Figure 10?

Discussion

The automated pharmacophore method implemented in DISCO has parallels to multiple regression or partial least squares analysis in QSAR. Most importantly, it examines the data to find all hypotheses that fit. The user then examines each hypothesis to see which best explains data not used in the derivation of the model. Just as in regression analysis, if more than one model remains, then it may be necessary to test additional molecules that will distinguish between the hypotheses. While in statistical analyses the models can be characterized by the R^2 and root-mean-square deviation, for DISCO the models are characterized by the number of points included, the tolerance of the match, and perhaps the energy of the conformers included or the union volume required for the model.

DISCO is a complement to 3D QSAR in that it (1) suggests possible pharmacophore maps and so automates the investigation of alignments possible within the data and (2) identifies molecules for which multiple conformers match the

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

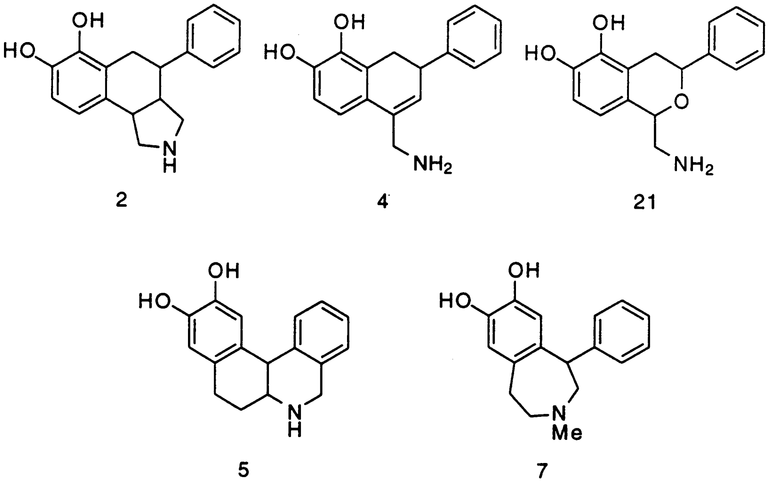

OH
HO
N
H
2
OH
HO
NH2
4
OH
HO
O
NH2
21
OH
HO
N
H
5
OH
HO
N
Me
7

Figure 5. The structures of D1 dopaminergic agonists considered in Table III. The number of distinct low-energy conformers is noted.

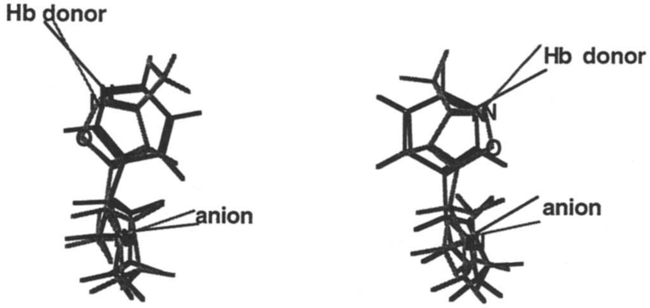

Hb donor
anion
Hb donor
N
anion

Figure 6. DISCO proposals of two different pharmacophore maps based on nicotine and ABT-418. All atoms are shown.

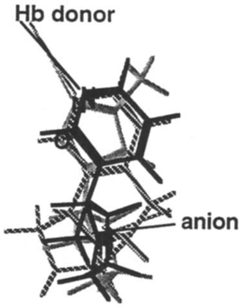

Hb donor
anion

Figure 7. The nicotinic agonist PHT (netted lines) fits only one of the pharmacophore maps (light lines). All atoms are shown.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

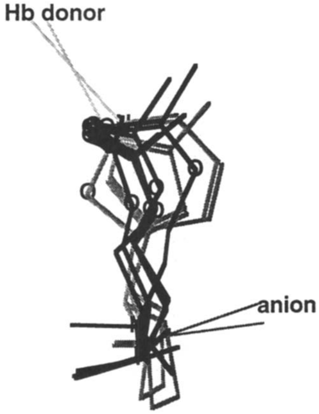


Figure 8. Superposition of several of the low-energy conformers of acetylcholine (heavy lines) over the hypothesis shown in Figure 6 (light lines). Only the heavy atoms are shown.

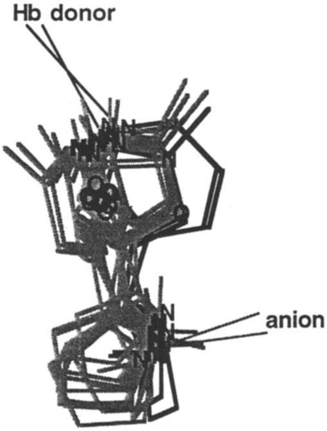


Figure 9. Superposition of low-energy conformers of a weaker agonist (light lines) over the hypothesis shown in Figure 6 (heavy lines). Only the heavy atoms are shown.

proposed pharmacophore map. In instances where many conformers of a molecule could be included in the 3D QSAR model, it may not be easy to decide which is correct. In this respect, the QSAR may aid the decision. If a 3D QSAR model is developed without such flexible molecules, then the proposed bioactive conformation of a molecule should be that conformation that the QSAR forecasts to be the most potent. Adding such conformations of flexible molecules back into the QSAR will show if the data are consistent.

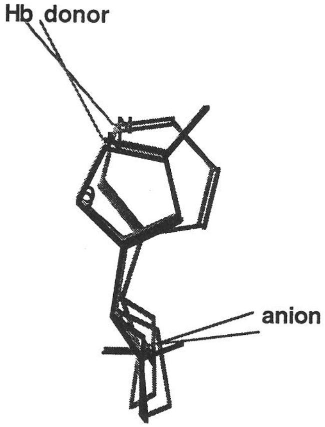


Figure 10. Superposition of a higher-energy conformer (1.85 kcal/mole) of the weaker agonist (heavy lines) over the hypothesis shown in Figure 6 (light lines). Only the heavy atoms are shown.

References

Allinger, M. L. *MMP2,* **1987,** Tripos Inc.; St. Louis MO.

Blaney, J. M., Crippen, G. M., Dearing, A. and Dixon, J. S. *DGEOM - Distance Geometry, QCP590,* **1990,** Quantum Chemistry Program Exchange, Indiana University; Bloomington, IN 47405.

Boyd, D. B. in *Reviews in Computational Chemistry;* Vol. 1; Lipkowitz, K. B. and Boyd, D. B., Eds. VCH: New York, 1990; Successes of Computer-Assisted Molecular Design, 355-371.

Brint, A. T. and Willett, P. *Journal of Chemical Information and Computer Sciences.* Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. **1987,** *27,* 152-158.

Bron, C. and Kerbosch, J. *Communications of the ACM.* Algorithm 457. Finding All Cliques of an Undirected Graph. **1973,** *16,* 575.

Cramer III, R. D., Patterson, D. E. and Bunce, J. D. *Journal of American Chemical Society.* Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. **1988,** *110,* 5959-5967.

Dewar, M. J. S., Zoebish, E. G., Healy, E. F. and Stewart, J. J. P. *Journal of the American Chemical Society.* AM1: A New General Purpose Quantum Mechanical Molecular Model. **1985,** *107,* 3902-3909.

Fujita, T. in *Drug Design: Fact or Fantasy?;* Vol. Jolles, G. and Wolldridge, K. R. H., Eds. Academic: London, 1984; The Role of QSAR in Drug Design, 19-33.

Garvey, D. S., Wasicak, J. T., Decker, M. W., Brioni, J. D., Buckley, M. J., Sullivan, J. P., Carrera, G. M., Holladay, M. W., Arneric, S. P. and Williams, M. *Journal of Medicinal Chemistry.* Novel Isoxazoles which Interact with Brain Cholinergic Channel Receptors Have Intrinsic Cognitive Enhancing and Anxiolytic Activities. **1994,** *37,* 1055-1059.

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Downloaded by UNIV MASSACHUSETTS AMHERST on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ch024

Goodford, P. J. *Journal of Medicinal Chemistry.* Drug Design by the Method of Receptor Fit. **1984,** *27,* 557-564.

Hansch, C. and Fujita, T. *Journal of the American Chemical Society.* Rho Sigma pi Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. **1964,** *86,* 1616-1626.

Hansch, C., Unger, S. H. and Forsythe, A. B. *Journal of Medicinal Chemistry.* Strategy in Drug Design. Cluster Analysis as an Aid in the Selection of Substituents. **1973,** *16,* 1212-1222.

Holladay, M. W., Wasicak, J. T., Donnelly-Roberts, D., Anderson, D. J., Pavlik, P., Martin, Y. C., Garvey, D. S., Sullivan, J. P. and Arneric, S. P.

Kim, K. H. in *3D QSAR in Drug Design. Theory Methods and Applications;* Vol. Kubinyi, H., Eds. ESCOM: Leiden, 1993; Comparison of Classical and 3D QSAR, 619-642.

Kubinyi, H., Eds. *3D QSAR in Drug Design. Theory Methods and Applications* **1993,** ESCOM; Leiden. pp. 759.

Langridge, R., Ferrin, T. E., Kuntz, I. D. and Connolly, M. L. *Science.* Real-Time Color Graphics in Studies of Molecular Interactions. **1981,** *211,* 661-667.

Marshall, G. R., Barry, C. D., Bosshard, H. E., Dammkoehler, R. A. and Dunn, D. A. in *Computer-Assisted Drug Design;* Vol. Olson, E. C. and Christoffersen, R. E., Eds. American Chemical Society: Washington, 1979; The Conformation Parameter in Drug Design: The Active Analog Approach, 205-226.

Martin, Y. C. *Quantitative Drug Design* **1978,** Dekker; New York. pp. 425.

Martin, Y. C. *Journal of Medicinal Chemistry.* A Practitioner's Perspective on the Role of Quantitative Structure-Activity Analysis in Medicinal Chemistry. **1981,** *24,* 229-237.

Martin, Y. C., Bures, M. G., Danaher, E. A., DeLazzer, J., Lico, I. and Pavlik, P. A. *Journal of Computer-Aided Molecular Design.* A Fast New Approach to Pharmacophore Mapping and its Application to Dopaminergic and Benzodiazepine Agonists. **1993,** *7,* 83-102.

Martin, Y. C., Kim, K.-H. and Lin, C. T. in *Linear Free Energy Relationships in Biology;* Vol. Charton, M., Eds. in press; Comparative Molecular Field Analysis: Comfa,

Martin, Y. C., Lin, C. T. and Wu, J. in *3D QSAR in Drug Design. Theory Methods and Applications;* Vol. Kubinyi, H., Eds. ESCOM: Leiden, 1993; Application of CoMFA to the Design and Structural Optimization of D1 Dopaminergic Agonists, 643-660.

Sheridan, R. P., Nilakantan, R., Dixon, J. S. and Venkataraghavan, R. *Journal of Medicinal Chemistry.* The Ensemble Approach to Distance Geometry: Application to the Nicotinic Pharmacophore. **1986,** *29,* 899-906.

RECEIVED May 1, 1995

Author Index

Akamatsu, Miki, 229,288
Andrea, Tariq A., 282
Brannigan, Lawrence H., 264
Charton, Marvin, 75
Chuman, Hiroshi, 171
Clark, Robert D., 264
Donovan, Stephen, 199
Draber, Wilfried, 186
Duewer, David L., 264
Fujita, Toshio, 1,13,36,229,288
Funaki, Yuji, 107
Gange, David M., 199
Hagiwara, Kenji, 213
Hansch, Corwin, 1,254
Hashimoto, Sho, 213
Henegar, Kevin, 199
Hermens, Joop L. M., 130
Hirano, H., 141
Hirono, S., 141
Hosaka, Hideo, 213
Ito, Atsushi, 171
Izumi, Kazuo, 107
Katagi, Toshiyuki, 48
Kim, Ki Hwan, 302
Kleier, D. A., 98
Kumazawa, Satoru, 171
Kurihara, Norio, 288
Kurita, Yasuyuki, 154
Leo, Albert J., 62
Liu, Q., 141
Lopata, Ronald J., 199
Magee, Philip S., 120
Martin, Yvonne Connolly, 318
Meki, Naoto, 154
Miyakado, Masakazu, 48,107
Moriguchi, I., 141
Nakagawa, Yoshiaki, 288
Nakayama, Akira, 213
Nishimura, Keiichiro, 288
Oettmeier, Walter, 186
Oikawa, Nobuhiro, 288
Parlow, John J., 264
Plummer, Ernest L., 240
Saishoji, Toshihide, 171
Schnur, Dora M., 264
Shimizu, Bun-ichi, 288
Takahashi, Masahiro, 107
Takano, Hirotaka, 107,154
Takayama, Chiyozo, 48,154
Tanaka, Shizuya, 48,107
Trebst, Achim, 186
Ueno, Tamio, 229,288
Verhaar, Henk J. M., 130
Yamagami, Chisako, 36

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix001

Affiliation Index

Abbott Laboratories, 302,318
American Cyanamid Company, 199
BIOSAR Research Project, 120
DuPont Agricultural Products, 98,282
FMC Corporation, 240
Fujitsu Kansai Systems Laboratory, 1
Kitasato University, 141
Kobe Pharmaceutical University, 36
Kureha Chemical Industry Company, Ltd., 171
Kyoto University, 13,36,229,288
Monsanto Company, 264
National Institute of Standards and Technology, 264
Nippon Soda Company, Ltd., 213
Pomona College, 1,62,254

In Classical and Three-Dimensional QSAR in Agrochemistry; Hansch, C., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1995.

Pratt Institute, 75
Ruhr-Universität, 186
Sumitomo Chemical Company, Ltd., 48,107,154
Tripos Inc., 264
University of California–San Francisco, 120
Utrecht University, 130
Zeria Pharmaceutical Company, Ltd., 141

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix001

Subject Index

A

- Acetylcholinesterase (AcChE) inhibition model, organophosphate end points, 138
- Acid dissociation constant, relationship to concentration factor and octanol–water partition coefficient, 102–103
- Acidity, relationship to activity in insecticidal uncouplers, 200,202–203
- Acridone(s), photosynthesis inhibition, 186–197
- Active conformations, computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines, 161–167
- Adaptive least squares
  - description, 141
  - use in agrochemical design, 156
- Advanced computer-aided chemistry system, antifungal tertiary amine analysis, 156–168
- Agrochemical(s)
  - requirements, 264
  - systemic translocation in plants, 107
- Agrochemical design
  - benzoylphenylurea larvicides, 17–19
  - bromobutide, 14–17
  - cyclic dicarboximide type fungicides, 19–21
  - EMIL, 24–31
  - QSAR, 23–24,31
  - synthetic pyrethroids, 22–23
  - use of computational procedures, 154
- Agrochemistry, role of classical QSAR, 2
- Allergens, model for allergic contact dermatitis, 120–128
- Allergic contact dermatitis model in modern pesticides
  - description, 122–127
  - immune system activation, 121–122
  - irritation ratings and categories of response, 127–128
  - modeling
    - allergens, 122–125
    - structural descriptors, 125
    - two-class regression analysis, 125–127
  - pesticide selection, 127
  - severity rating of dermal effect, 121
- Amines, antifungal tertiary, computer-aided molecular modeling and structure–activity analyses, 154–168
- Amino acid side chains applied to quantitative structure–activity analyses of oligopeptides, 229–238
- Aniline(s), radical toxicity, 258–260
- Aniline mustard antitumor agents, QSAR, 255–256,260–261
- Antifungal amine compounds
  - design, 156–157
  - synthesis, 157–159
- Antifungal tertiary amines, computer-aided molecular modeling and structure–activity analyses, 154–168
- Arg-Gly-Asp peptides, platelet aggregation inhibition, 220–233,238
- Azole-type fungicides, classification, 171,172*f*
- Azolylmethylcyclopentanols
  - antifungal activities, 173–176
  - binding mode to cytochrome P450, 183–184
  - conformational analysis procedure, 177
  - experimental description, 172

Pratt Institute, 75
Ruhr-Universität, 186
Sumitomo Chemical Company, Ltd., 48,107,154
Tripos Inc., 264
University of California–San Francisco, 120
Utrecht University, 130
Zeria Pharmaceutical Company, Ltd., 141

Subject Index

A

Acetylcholinesterase (AcChE) inhibition model, organophosphate end points, 138
Acid dissociation constant, relationship to concentration factor and octanol–water partition coefficient, 102–103
Acidity, relationship to activity in insecticidal uncouplers, 200,202–203
Acridone(s), photosynthesis inhibition, 186–197
Active conformations, computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines, 161–167
Adaptive least squares
  description, 141
  use in agrochemical design, 156
Advanced computer-aided chemistry system, antifungal tertiary amine analysis, 156–168
Agrochemical(s)
  requirements, 264
  systemic translocation in plants, 107
Agrochemical design
  benzoylphenylurea larvicides, 17–19
  bromobutide, 14–17
  cyclic dicarboximide type fungicides, 19–21
  EMIL, 24–31
  QSAR, 23–24,31
  synthetic pyrethroids, 22–23
  use of computational procedures, 154
Agrochemistry, role of classical QSAR, 2
Allergens, model for allergic contact dermatitis, 120–128
Allergic contact dermatitis model in modern pesticides
  description, 122–127
  immune system activation, 121–122
  irritation ratings and categories of response, 127–128
  modeling
    allergens, 122–125
    structural descriptors, 125
    two-class regression analysis, 125–127
  pesticide selection, 127
  severity rating of dermal effect, 121
Amines, antifungal tertiary, computer-aided molecular modeling and structure–activity analyses, 154–168
Amino acid side chains applied to quantitative structure–activity analyses of oligopeptides, 229–238
Aniline(s), radical toxicity, 258–260
Aniline mustard antitumor agents, QSAR, 255–256,260–261
Antifungal amine compounds
  design, 156–157
  synthesis, 157–159
Antifungal tertiary amines, computer-aided molecular modeling and structure–activity analyses, 154–168
Arg-Gly-Asp peptides, platelet aggregation inhibition, 220–233,238
Azole-type fungicides, classification, 171,172*f*
Azolylmethylcyclopentanols
  antifungal activities, 173–176
  binding mode to cytochrome P450, 183–184
  conformational analysis procedure, 177
  experimental description, 172

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Azolylmethylcyclopentanols—*Continued*
  partition coefficient, 176
  QSAR, 177–178,183
  steric parameter determination, 174–176
  structures, 171,172*f*
  superposition procedure, 177
  synthetic scheme, 173
  three-dimensional shape comparison, 178–181
  X-ray crystallographic and conformational analysis, 181–183

B

Benzoic acids, dissociation constants, 306–308
1,4-Benzoquinones, photosynthesis inhibition, 186–197
Benzoylphenylurea larvicides, design, 17–19
Bifenthrin, discovery using QSAR paradigm, 241
Bioactive compounds, design strategy, 13
Bioactivity, role of hydrophobicity, 5–6
Biochemical expert system, description, 156
Bioconcentration model, organophosphate end points, 135,137–138
Biodegradation, noncongeneric structure–toxicity correlation using fuzzy adaptive least squares, 148,150–151
Bioisosteric structural transformation
  identification of rules, 25–30
  system operation, 30–31
Biological activity and toxicity data, description using activity ratings, 141
Biological quantitative structure–activity relationships, lateral validation, 303
Bleaching herbicidal compounds, hydrophobicity–systemic activity relationship, 113–118
Bromobutide, design, 14–17
Bromoxynil, irritation responses, 121

C

C-QSAR system, description, 156
Capacity factors, heteroaromatic compounds, 36–47
Captan, irritation responses, 121
Carbamate-based pesticides, hydrophobicity calculation by computer, 62–68
Carcinogenicity in rodents, noncongeneric structure–toxicity correlation using fuzzy adaptive least squares, 147–150
Central composite design, use of QSAR paradigm, 243–250
Charge-transfer parameters, intermolecular force dependence, 77
Chemical structure, relationship to phloem mobility, 99
Chloroform–water partition coefficient, heteroaromatic compounds, 36–47
Cholinesterase inhibitors, QSAR, 257–258,261
Classical quantitative structure–activity relationship
  advantages, 6–7,318–319
  benzoylphenylurea larvicide, 17–19
  bromobutide design, 14–17
  collinearity problem, 4
  comparison
    to comparative molecular field analysis, 8–9
    to EMIL, 31
  description, 23–24
  effect of hydrophobicity on bioactivity, 5–6
  equivocal definition of parameters, 5
  expansion, 7
  lateral validation, 303–305
  limitations, 319
  postulates, 2
  quality, 2–4
  substituent selection, 4–5
  underparameterization, 5
  use in agrochemical design, 2,14
Collinearity problem, classical QSAR, 4
Commercial compounds, discovery using QSAR paradigm, 241
Comparative molecular field analysis
  advantages, 9–10,302
  analytical steps, 305
  comparison to classical QSAR, 8–9
  covalent bond formation factor, 7
  function, 264
  lateral validation, 305–313

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Comparative molecular field analysis—*Continued*
  procedure, 7–8
  two-component model, 305–306
Compounds derived from partitioning systems, hydrophobicity parameter, 36–47
Computational procedures, design of agrochemicals and medicines, 154
Computer-aided molecular design, objective, 318
Computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines, 154–168
  active conformation, 161–167
  design systems, 155–157
  experimental description, 154
  molecular orbital calculations, 161
  pK_a estimation, 159–161
  sterol analysis, 159
  superimposition of three-dimensional structures, 166–168
  synthesis, 157–159,160*t*
Concentration factor, relationship to acid dissociation constant and octanol–water partition coefficient, 102–103
Contact dermatitis model in modern pesticides, Allergic, *See* Allergic contact dermatitis model in modern pesticides
Cyclic dicarboximide type fungicides, design, 19–21

D

Database-aided bioisosteric structural transformation procedure and QSAR for agrochemical design, 13–31
DDT analogues, hydrophobicity calculation by computer, 69–70
Delayed-type contact dermatitis model in modern pesticides, *See* Allergic contact dermatitis model in modern pesticides
Dermal effects, severity rating, 121
Design, antifungal tertiary amines, 155–157
Design project management, 250–252
Dibenzoylhydrazine
  insecticidal activity, 288–289
  structure, 289
  three-dimensional QSAR, 289–301
Dibenzoylhydrazine-type ecdysone agonists, three-dimensional QSAR, 288–301
Diflufenican, hydrophobicity–bleaching activity relationship, 115–118
Difunon, hydrophobicity–bleaching activity relationship, 115–118
Diniconazole, hydrophobicity–systemic activity relationship, 108–113
2,4-Dinitrochlorobenzene, irritation responses, 121
Dioxapyrrolomycin, insecticidal activity, 199–200
Dipole moment, intermolecular force dependence, 75–94
DISCO strategy for pharmacophore mapping
  comparison to three-dimensional QSAR, 325,328
  evaluation of results, 324–326*f*
  finding of superpositions and conformers that match in three dimensions, 322,324
  generation of all low-energy conformers as potentially bioactive, 320,323*f*
  identification of set of active compounds on which to base pharmacophore, 320,321*f*
  labeling of charged, hydrogen bonding, and hydrophobic points that may form part of superposition rule, 322,323*f*
  nicotinic agonist comparison, 325–328*f*
Discovery programs, application of QSAR paradigm, 240–252
Dissociation constants, benzoic acids and phenylacetic acids, 306–308
DNA reactions, QSAR, 255

E

Ecdysone agonists, three-dimensional QSAR, 288–301
Electron-donating effect, role in enzyme inhibitory activity, 282–287

Electron-withdrawing effects, role in enzyme inhibitory activity, 282–287
Electronic factor, role in classical QSAR, 2
Environmental toxicology and chemistry, QSAR, 130–138
Ergosterol
  biosynthesis inhibitor design, 157–160*t,f*
  biosynthesis pathway, 156–159
Example-mediated innovation for lead evolution (EMIL)
  bromobutide design, 14–17
  comparison to QSAR, 31
  cyclic dicarboximide-type fungicide design, 19–21
  description, 14,24–25,156
  identification of bioisosteric transformation rules, 25–30
  operation of bioisosteric transformation system, 30–31
  use in agrochemical design, 14

F

Fecosterol, basis for design of new fungicides, 154–168
Fenpropidin, structure, 156
Fenpropimorph, structure, 156
Fenpyroximate, potential activity testing, 245–250
Fish lethal concentration quantitative structure–activity relationship, 131–134
Fluridone, hydrophobicity–bleaching activity relationship, 115–118
Flurtamone, hydrophobicity–bleaching activity relationship, 115–118
Folate synthetase inhibitors, QSAR, 256–257,260
Fungicidal azolylmethylcyclopentanols, QSAR and three-dimensional shape studies, 171–184
Fungicidal triazole compounds, hydrophobicity–systemic activity relationship, 107–114
Fungicides
  cyclic dicarboximide-type, design, 19–21
  1,2,4-thiadiazoline-type, QSAR and molecular modeling, 213–227
Fuzzy adaptive least squares
  applications, 141–142
  noncongeneric structure–toxicity correlation, 142–151

G

Geraniol, description, 123
Gluten exorphin analogues, opioid activity, 223–238
Graphical analysis, pyrrole insecticides, 205–208

H

Hansch–Fujita model
  explanation, 75–94
  transport parameter dependence on intermolecular forces, 91–94
  use in agrochemical design, 156
Hansch–Fujita quantitative structure–activity relationship, *See* Classical quantitative structure–activity relationship
Hazards of pesticides, testing, 120
Herbicide(s)
  bleaching, hydrophobicity–systemic activity relationship, 113–118
  light-dependent, structure–activity analyses, 220–227
  QSAR, scaled rank sum statistics, 264–280
  1,2,4-thiadiazoline-type, QSAR and molecular modeling, 213–227
Heteroaromatic compounds derived from partitioning systems, hydrophobicity parameter, 36–47
Hydrogen-acceptor scale, description, 45–47
Hydrogen-bond parameterization, intermolecular force equation, 84
Hydrogen bonding, transport parameter dependence, 91–92
Hydrogen bonding parameter, intermolecular force dependence, 77

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Hydrophobic interactions
  organic compounds, 254
  unimportance in QSAR, 254–261
Hydrophobicity
  carbamate-based pesticides, 62–68
  DDT analogues, 69–70
  importance in biological activity, 36
  of pesticides, calculation by computer, 62–74
  phosphorus-containing pesticides, 64,69
  role in bioactivity, 5–6
  role in classical QSAR, 2
  urea-type herbicides, 70–74
  See also Octanol–water partition coefficient
Hydrophobicity parameter
  amino acid side chains applied to quantitative structure–activity analyses of oligopeptides
    comparison to other hydrophobicity indexes, 229
    components, 230
    opioid activity of gluten exorphin analogues, 233–238
    platelet aggregation inhibition of Arg-Gly-Asp peptides, 230–233,238
  heteroaromatic compounds derived from partitioning systems, 36–47
    estimation, 36–37
    experimental procedure, 37–38
    hydrogen-acceptor scale, 45–47
    octanol–water partition coefficient–capacity factor relationship, 39–45
    octanol–water partition coefficient–chloroform–water partition coefficient relationship, 38–39
  previous studies, 229
  role in enzyme inhibitory activity, 282–287
Hydrophobicity–systemic activity relationship, 107–118
  bleaching herbicidal compounds
    bleaching patterns, 115–117
    distribution patterns according to octanol–water partition coefficient, 113–114
Hydrophobicity–systemic activity relationship—*Continued*
  bleaching herbicidal compounds—*Continued*
    hydrophobicity–bleaching activity relationship, 117–118
    structures, 115
  fungicidal triazole compounds
    distribution patterns according to octanol–water partition coefficient, 113–114
    hydrophobicity–systemic activity relationship, 110–113
    structures, 108
    systemic activity, 109–111*t*
    translocation in cucumber, 109,110*f*
20-Hydroxyecdysone
  insecticidal activity, 288–289
  structure, 288
  three-dimensional QSAR, 289–301

I

Immune system activation, allergic contact dermatitis model, 121–122
Inhibitors, photosystem II, structure–activity relationships, 186–197
Insecticidal uncouplers, QSAR, 199–211
  acidity vs. activity, 200,202–203
  design, 211
  discovery of uncoupling, 200
  lipophilicity vs. activity, 200,202–203
  mode of action, 200,201*f*
  pyrrole insecticides, 203–211
  steric effect of substituents vs. activity, 202–203
  structures, 203
Intermolecular force equation
  description, 76–77,94
  dipole moment parameterization, 80–83
  hydrogen-bond parameterization, 84
  ionic group parameterization, 84–85
  steric effect parameterization, 83–84
Intermolecular force model, quantitative description of transport parameters, 75–94

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Ionic group parameterization, intermolecular force equation, 84–85
Ipconazole, QSAR and three-dimensional shape studies, 171–184
Isothiocyanates, QSAR, 256,260

L

Larvicides, benzoylphenylurea, design, 17–19
Lateral validation
  application in combining multiple equations, 313–315
  biological QSAR, classical, 303
  comparative molecular field analysis
    dissociation constants of benzoic acids and phenylacetic acids, 306–308
    rate constant for elimination of 2-Z-phenylethyl *p*-Y-arenesulfonates with potassium *tert*-butoxide in *tert*-butyl alcohol, 308–313
  data base, 302–303
  description, 302
  future work, 317
  lattice point position effect on coefficients, 315–317
  physical organic chemistry QSAR, 303–305
Lattice point position, role on coefficients, 315–317
Light-dependent herbicides, structure–activity analyses, 220–227
Lipophilicity, relationship to activity in insecticidal uncouplers, 200,202–203
Log *P*, *See* Octanol–water partition coefficient

M

MACCS II system, description, 156
Mathematical models, QSAR, 98–105
Medicines, design using computational procedures, 154
Metconazole, QSAR and three-dimensional shape studies, 171–184
Methoxyureas, hydrophobicity calculation by computer, 70–71
Molar refractivity, role in enzyme inhibitory activity, 282–287
Molecular field analysis, comparison to scaled rank sum statistics in herbicide quantitative structure–activity relationship, 274–279,280
Molecular geometries, organophosphorus pesticides, 49,51–56
Molecular modeling
  antifungal tertiary amines, computer aided, *See* Computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines
  fungicides and herbicides having 1,2,4-thiadiazoline structures, 213–227
Molecular orbits, computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines, 161
Molting hormone activity, measurement, 289
Multiple equations, use of lateral validation in combining, 313–315
Multiple regression analysis
  comparison to scaled rank sum statistics in herbicide QSAR, 271–273*t*,280
  octanol–water partition coefficient for organophosphorus pesticides, 54,57–58
Multivariate approaches using quantum-chemical parameters, 133–138

N

1,4-Naphthoquinones, photosynthesis inhibition, 186–197
Nerol, description, 123
Neural network models of QSAR, 282–287
  comparison to regression QSAR, 287
  enzyme inhibitory activities, 283,284*t*
  experimental procedure, 283–284
  factors affecting enzyme inhibitory activity, 284–285
  network architecture, 283
  structures, 283,284*t*
  substituent charge effect on enzyme activity inhibition, 286–287

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Nicotinic agonist, comparison using DISCO strategy for pharmacophore mapping, 325–328*f*
Nonallergens, model for allergic contact dermatitis, 120–128
Noncongeneric structure–toxicity correlation using fuzzy adaptive least squares, 141–152
  advantages, 151
  aquatic toxicity of organic chemicals, 144–147
  biodegradation, 148,150–151
  carcinogenicity in rodents, 147–150
  method, 142–144
Norflurazon, hydrophobicity–bleaching activity relationship, 115–118

O

Octanol–water partition coefficient
  heteroaromatic compounds, 36–47
  organophosphorus pesticides
    calculated vs. experimental values, 58–60
    calculation procedure, 49
    compounds, 49–54*t*
    determination methods, 48–49
    molecular geometries, 49,51–56
    multiple regression analysis, 54,57–58
  relationship to concentration factor and acid dissociation constant, 102–103
  theoretical estimation, 48–60
  See also Hydrophobicity
Oligopeptides, hydrophobicity parameter of amino acid side chains applied to QSAR analyses, 229–239
Opioid activity, gluten exorphin analogues, 233–238
Optimization of lead molecules, use of QSAR paradigm, 241
ORACLE system, description, 156
Organic chemicals, noncongeneric structure–toxicity correlation using fuzzy adaptive least squares, 144–147
Organic compounds, hydrophobic interactions, 254
Organophosphate end points of quantitative structure–activity relationships
  AchE inhibition model, 138
  bioconcentration model, 135,137–138
  determination processes, 135,137*f*
Organophosphorus pesticides, octanol–water partition coefficients, 48–60

P

Parameters, definition, 5
Partition coefficient, *See* Hydrophobicity, Octanol–water partition coefficient
Partitioning systems, various, hydrophobicity parameters of heteroaromatic compounds, 36–47
Peptidic and peptidomimetic bioactive compounds, interest, 229
Pesticides
  allergic contact dermatitis model, 120–128
  hydrophobicity calculated by computer, 62–74
  QSAR from mathematical models for systemic behavior, 98–105
Pharmacophore mapping, DISCO strategy, 320–328
Phenols, radical toxicity, 258–260
Phenylacetic acids, dissociation constants, 306–308
Phenylproparginols, scaled rank sum statistics in herbicide QSAR, 269–273*t*
Phloem mobility, relationship to chemical structure, 99
Phloem systemic compounds, function, 98
Phloem systemic nematocides, design, 104–105
Phloem systemic pesticide, advantages, 98–99
Phosphorus-containing pesticides, hydrophobicity calculation, 64,69
Photosystem II inhibitors, structure–activity relationships, 186–197
Photosystem II reaction center
  amino acid sequence in binding, 186–188*f*
  composition, 186
  inhibitor mechanism, 186
  types of inhibitor families, 186–187

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Phthalic anhydride, irritation responses, 121
PhX transport parameters, intermolecular force dependence, 78–80,82
Physical organic chemistry structure–activity relationships, lateral validation, 303–305
pK_a, model development for pyrrole insecticides, 205,206*f*
pK_a estimation, computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines, 159–161
Platelet aggregation inhibition, Arg-Gly-Asp peptides, 230–233,238
Polarizability parameter, intermolecular force dependence, 77
Postulates, classical QSAR, 2
Pyrazole olefinic nitriles, scaled rank sum statistics in herbicide QSAR, 272–279
Pyrethroids, synthetic, design, 22–23
Pyrrole insecticidal analysis
  compounds, 204
  graphical analysis, 205–208
  lethal concentration, 204
  octanol–water partition coefficient values, 204
  pK_a model development, 205,206*f*
  pK_a values, 204
  QSAR, 207*f*,208–211
Pyrroledicarboxylates, hydrophobicity calculation by computer, 71,74*f*

Q

Quality, classical QSAR, 2–4
Quantitative structure–activity relationships (QSAR)
  acridone photosystem II inhibitors, 186–197
  aniline mustard antitumor agents, 255–256,260–261
  cholinesterase inhibitors, 257–258,261
  comparative, importance of hydrophobic interactions, 254–261
  development, 99
  dibenzoylhydrazine-type ecdysone agonists, 288–301
  DNA reactions, 255
Quantitative structure–activity relationships (QSAR)—*Continued*
  environmental toxicology and chemistry
    advantages, 138
    chemical reactivity, 133–135
    experimental description, 131
    limitations, 131–134
    multivariate approaches using quantum-chemical parameters, 133–138
  folate synthetase inhibitors, 256–257,260
  fungicidal azolylmethylcyclopentanols, 171–184
  fungicides and herbicides having 1,2,4-thiadiazoline structures, 213–227
  insecticidal uncouplers, 199–211
  isothiocyanates, 26,260
  mathematical models for systemic behavior of pesticides, 98–105
    concentration factor, dependence on acid dissociation constant and octanol–water partition coefficient, 102–103
    description of model, 99–102
    experimental description, 99
    propesticide approach to design of phloem systemic nematocides, 104–105
    rules for rendering xenobiotics phloem systems, 103–104
  oligopeptides, 229–239
  pyrrole insecticides, 207–211
  quinone photosystem II inhibitors, 186–197
  reasons for interest, 264
  scaled rank-sum statistics, 264–280
  steroidal ecdysone agonists, 288–301
  three-dimensional, *See* Three-dimensional quantitative structure–activity relationships
Quantitative structure–activity relationship (QSAR) paradigm, 240–252
  design project management, 250–252
  development, 240
  discovery of commercial compounds, 241
  efficient optimization of lead molecules
    central composite design, 243–250
    sequential simplex optimization, 242–243
  new lead generation, 241
  success criteria, 240–241

Quinone photosystem II inhibitors, structure–activity relationships, 186–197

R

Radish seedlings, bleaching patterns, 115–117
Rank transformation, limitations, 265
Rate constant for elimination of 2-Z-phenylethyl *p*-Y-arenesulfonates with potassium *tert*-butoxide in *tert*-butyl alcohol, 308–313
Regression quantitative structure–activity relationship
  biological activity–physicochemical property relationship, 283
  comparison to neural network models of QSAR, 287

S

Scaled rank sum(s), applications, 265
Scaled rank-sum statistics in herbicide quantitative structure–activity relationship
  advantages, 280
  calculations, 268–270*f*
  comparison
    to comparative molecular field analysis, 274–279,280
    to multiple regression analysis, 271–273*t*,280
  experimental description, 265
  phenylproparaginols, 269–272,273*t*
  precision, 274,275*f*
  pyrazole olefinic nitriles, 272–279
  simulated titrations, 266–268,270*f*
Sequential simplex optimization, use of QSAR paradigm, 242–243
Shape studies, three-dimensional, fungicidal azolylmethylcyclopentanols, 171–184
Statistical predictive expertizing in chemical and technological refined experimentation, description, 156
Steric effect of substituents, relationship to activity in insecticidal uncouplers, 202–203
Steric effect parameter(s), intermolecular force dependence, 77
Steric effect parameterization, intermolecular force equation, 83–84
Steric factor, role in classical quantitative structure–activity relationship, 2
Steroidal agonists, three-dimensional QSAR, 288–301
Sterol, computer-aided molecular modeling and structure–activity analyses, 159
Structure–activity analyses
  antifungal tertiary amines, *See* Computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines
  oligopeptides, hydrophobicity parameter of amino acid side chains, 229–239
  quinone and acridone photosystem II inhibitors
    data base, 190,192–193
    experimental procedure, 187
    inhibitory potency, 188–192
    previous studies, 187
    QSAR, 192,194–196
    regression equations, 196
  Δ^2–1,2,4-thiadiazolines and light-dependent herbicides
    electrostatic similarity, 224–226*f*
    herbicidal activity, 220–221
    mode of action, 222
    molecular electrostatic potential and conformation, 222–224
    steric similarity, 225,226*t*
    structural requirements, 225,227
    structures, 220,221*f*
  Δ^3-1,2,4-thiadiazolines
    chemical reactivity, 215,217*t*
    mode of action, 215
    molecular orbital parameters, 215–217
    pesticidal activities, 213
    preventive activity, 214
    QSAR, 217–220
    structure, 214
Structure–toxicity correlation using fuzzy adaptive least squares, noncongeneric, *See* Noncongeneric structure–toxicity correlation using fuzzy adaptive least squares

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Substituent selection, classical QSAR, 4–5
Sulfonylureas, hydrophobicity calculation by computer, 71,72*f*
Swain–Lupton's electronic parameters, role in enzyme inhibitory activity, 282–287
Synthesis, computer-aided molecular modeling and structure–activity analyses of antifungal tertiary amines, 157–160*t*
Synthetic pyrethroids, design, 22–23
Systemic activity–hydrophobicity relationship, *See* Hydrophobicity–systemic activity relationship
Systemic behavior of pesticides, QSAR from mathematical models, 98–105
Systemic translocation, problem in agrochemicals in plants, 107

T

Tebufenozide
  insecticidal activity, 288–289
  structure, 289
  three-dimensional QSAR, 289–301
Tertiary amines, antifungal, computer-aided molecular modeling and structure–activity analyses, 154–168
1,2,4-Thiadiazolines
  pesticidal activity, 213
  QSAR and molecular modeling, 214–227
Δ^2-1,2,4-Thiadiazolines and light-dependent herbicides, structure–activity analyses, 220–227
Δ^3-1,2,4-Thiadiazolines, structure–activity analyses, 213–220
Three-dimensional quantitative structure–activity relationship
  comparative molecular field analysis, 7, 292
  comparison to DISCO strategy for pharmacophore mapping, 325,328
  dibenzoylhydrazine(s), 292–294,299
  dibenzoylhydrazine–steroidal ecdysone combined set comparative molecular field analysis, 294–296,299
  experimental materials, 289–291
  future work, 301
  getting started, 7
Three-dimensional quantitative structure–activity relationship—*Continued*
  limitations, 319
  molecular modeling procedure, 291–292
  strengths, 319
Three-dimensional shape studies, fungicidal azolylmethylcyclopentanols, 171–184
Toxicity–structure correlation using fuzzy adaptive least squares, noncongeneric, *See* Noncongeneric structure–toxicity correlation using fuzzy adaptive least squares
Toxicology, QSAR, 130–138
Transport parameter, dependence on intermolecular forces, 75–94
  charge-transfer parameters, 77
  combination
    of data sets for skeletal groups with same hybridization, 85–86
    of data sets studied in different media, 86–88
  composition of substituent effect, 78
  contributions from individual intermolecular forces, 94
  dipole moment parameterization in intermolecular force equation, 80–83
  experimental description, 77
  Hansch–Fujita model, 92–94
  hydrogen-bond parameterization in intermolecular force equation, 84
  hydrogen bonding, 77,92
  influencing factors, 89–90
  intermolecular force(s) and quantities upon which they depend, 75–76
  intermolecular force equation, 76–77,94
  ionic group parameterization in intermolecular force equation, 84–85
  nature, 92
  parameter contributions, 91–92
  PhX transport parameters, 78–80,82
  polarizability parameter, 77
  steric effect parameter(s), 77
  steric effect parameterization in intermolecular force equation, 83–84
  test set requirements, 78
sym-Triazines, hydrophobicity calculation by computer, 71,73*f*

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Triazole compounds, hydrophobicity–systemic activity relationship, 107–114
1,1,1-Trichloro-2,2-bis(*p*-chlorophenyl)-ethane analogues, hydrophobicity calculation by computer, 69–70
Two-class regression analysis, allergic contact dermatitis model, 125–127

U

Uncouplers, insecticidal, *See* Insecticidal uncouplers
Uncoupling, 200–203
Uniconazole, hydrophobicity–systemic activity relationship, 108–113
Urea-type herbicides, hydrophobicity calculation by computer, 70–73

X

Xenobiotics, phloem systemic rules, 103–104

Downloaded by 89.163.34.136 on October 14, 2012 | http://pubs.acs.org
Publication Date: May 5, 1995 | doi: 10.1021/bk-1995-0606.ix002

Production: Amie Jackowski
Indexing: Deborah H. Steiner
Acquisition: Anne Wilson & Barbara E. Pralle
Cover design: Cornithia Allen Harris

Printed and bound by Maple Press, York, PA